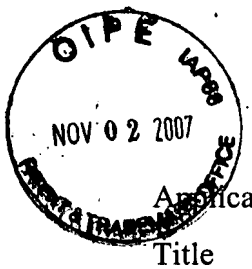
11 05 07

DPW



PATENT
674575-2004

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	Gutteridge, et al.
Title	:	ADHESION MOLECULES
Application No.	:	10/615,515
Filing Date	:	July 8, 2003
Examiner	:	Rosanne Kosson
Art Unit	:	1651
Confirmation No.	:	9209

745 Fifth Avenue,
New York, NY 10151

EXPRESS MAIL

Mailing Label Number: EV 845293456

Date of Deposit: November 2, 2007

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" Service under 37 CFR 1.10 on the date indicated above and is addressed to: **Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

SALLY ANN Homer
(Typed or printed name of person mailing paper or fee)

SAHomer
(Signature of person mailing paper or fee)

COMMUNICATION FORWARDING PRIORITY DOCUMENT

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

Enclosed is a certified copy of the priority document for the above-identified application. Applicants hereby claim priority under 35 U.S.C. §§ 119 and/or 120 from GB Application No. PCT/GB2002/00107.

Acknowledgment of the claim of priority and receipt of this certified copy is requested.

Respectfully submitted,

FROMMER LAWRENCE & HAUG LLP
Attorneys for Applicants

By: Thomas J. Kowalski by Angela M. Collison
Thomas J. Kowalski
Reg. No. 32,147

Angela M. Collison
Reg. No. 51,107
(212) 588-0800

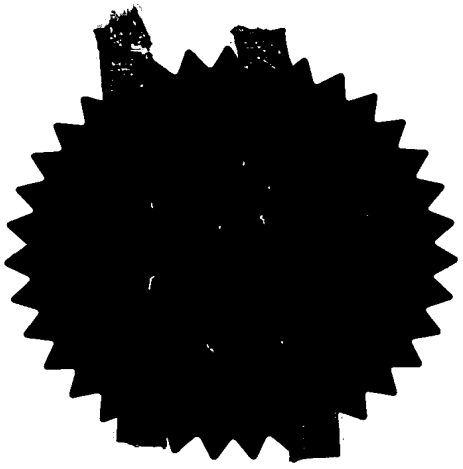
The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation and Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General hereby certify that annexed hereto is a true copy of the international application filed on 11 January 2002 under the Patent Cooperation Treaty at the UK Receiving Office. The application was allocated the number PCT/GB2002/000107

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or the inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed; M E Taylor

Dated; 06 November 2006

HOME

PCT

REQUEST

The undersigned requests that the present international application be processed according to the Patent Cooperation Treaty.

For receiving Office use only

PCT/GB 02 / 00107
International Application No.

11 JANUARY 2002 11/01/02
International Filing Date

United Kingdom Patent Office
PCT International Application
Name of receiving Office and "PCT International Application"

Applicant's or agent's file reference *(if desired) (12 characters maximum)* P026477WO

Box No. I TITLE OF INVENTION
ADHESION MOLECULES

Box No. II APPLICANT ☐ This person is also inventor

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

INPHARMATICA LIMITED

60 Charlotte Street
London
W1T 2NU
GB

Telephone No.

Facsimile No.

Teleprinter No.

Applicant's registration No. with the Office

State *(that is, country)* of nationality: GB

State *(that is, country)* of residence: GB

This person is applicant for the purposes of: ☐ all designated States ☒ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

GUTTERIDGE, Alex
INPHARMATICA LIMITED
60 Charlotte Street
London
W1T 2NU
GB

This person is:
☐ applicant only
☒ applicant and inventor
☐ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality: GB

State *(that is, country)* of residence: GB

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☒ the United States of America only ☐ the States indicated in the Supplemental Box

☒ Further applicants and/or (further) inventors are indicated on a continuation sheet.

Box No. IV AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE

The person identified below is hereby/has been appointed to act on behalf of the applicant(s) before the competent International Authorities as: ☒ agent ☐ common representative

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

MERCER, Christopher Paul

CARPMAELS & RANSFORD
43 BLOOMSBURY SQUARE
LONDON WC1A 2RA
GB

Telephone No.
020 7242-8692

Facsimile No.
020 7405 4166

Teleprinter No.

Agent's registration No. with the Office

☐ **Address for correspondence:** Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.

Sheet No. ...2...

Continuation of Box No. III **FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)**

If none of the following sub-boxes is used, this sheet should not be included in the request.

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

FAGAN, Richard Joseph
INPHARMATICA LIMITED
60 Charlotte Street
London
W1T 2NU
GB

This person is:
- ☐ applicant only
- ☒ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office:

State *(that is, country)* of nationality: US

State *(that is, country)* of residence: GB

This person is applicant for the purposes of:
- ☐ all designated States
- ☐ all designated States except the United States of America
- ☒ the United States of America only
- ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

PHELPS, Christopher Benjamin
INPHARMATICA LIMITED
60 Charlotte Street
London
W1T 2NU
GB

This person is:
- ☐ applicant only
- ☒ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office:

State *(that is, country)* of nationality: GB

State *(that is, country)* of residence: GB

This person is applicant for the purposes of:
- ☐ all designated States
- ☐ all designated States except the United States of America
- ☒ the United States of America only
- ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

This person is:
- ☐ applicant only
- ☐ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office:

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of:
- ☐ all designated States
- ☐ all designated States except the United States of America
- ☐ the United States of America only
- ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

This person is:
- ☐ applicant only
- ☐ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office:

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of:
- ☐ all designated States
- ☐ all designated States except the United States of America
- ☐ the United States of America only
- ☐ the States indicated in the Supplemental Box

☐ Further applicants and/or (further) inventors are indicated on another continuation sheet.

Box No. V DESIGNATION OF STATES *Mark the applicable check-boxes below; at least one must be marked.*

The following designations are hereby made under Rule 4.9(a):

Regional Patent

☒ AP **ARIPO Patent:** GH Ghana, GM Gambia, KE Kenya, LS Lesotho, MW Malawi, MZ Mozambique, SD Sudan, SL Sierra Leone, SZ Swaziland, TZ United Republic of Tanzania, UG Uganda, ZM Zambia, ZW Zimbabwe, and any other State which is a Contracting State of the Harare Protocol and of the PCT *(if other kind of protection or treatment desired, specify on dotted line)* .....

☒ EA **Eurasian Patent:** AM Armenia, AZ Azerbaijan, BY Belarus, KG Kyrgyzstan, KZ Kazakhstan, MD Republic of Moldova, RU Russian Federation, TJ Tajikistan, TM Turkmenistan, and any other State which is a Contracting State of the Eurasian Patent Convention and of the PCT

☒ EP **European Patent:** AT Austria, BE Belgium, CH & LI Switzerland and Liechtenstein, CY Cyprus, DE Germany, DK Denmark, ES Spain, FI Finland, FR France, GB United Kingdom, GR Greece, IE Ireland, IT Italy, LU Luxembourg, MC Monaco, NL Netherlands, PT Portugal, SE Sweden, TR Turkey, and any other State which is a Contracting State of the European Patent Convention and of the PCT

☒ OA **OAPI Patent:** BF Burkina Faso, BJ Benin, CF Central African Republic, CG Congo, CI Côte d'Ivoire, CM Cameroon, GA Gabon, GN Guinea, GQ Equatorial Guinea, GW Guinea-Bissau, ML Mali, MR Mauritania, NE Niger, SN Senegal, TD Chad, TG Togo, and any other State which is a member State of OAPI and a Contracting State of the PCT *(if other kind of protection or treatment desired, specify on dotted line)* .....

National Patent *(if other kind of protection or treatment desired, specify on dotted line)*:

☒ AE United Arab Emirates
☒ AG Antigua and Barbuda
☒ AL Albania
☒ AM Armenia
☒ AT Austria
☒ AU Australia
☒ AZ Azerbaijan
☒ BA Bosnia and Herzegovina
☒ BB Barbados
☒ BG Bulgaria
☒ BR Brazil
☒ BY Belarus
☒ BZ Belize
☒ CA Canada
☒ CH & LI Switzerland and Liechtenstein
☒ CN China
☒ CO Colombia
☒ CR Costa Rica
☒ CU Cuba
☒ CZ Czech Republic
☒ DE Germany
☒ DK Denmark
☒ DM Dominica
☒ DZ Algeria
☒ EC Ecuador
☒ EE Estonia
☒ ES Spain
☒ FI Finland
☒ GB United Kingdom
☒ GD Grenada
☒ GE Georgia
☒ GH Ghana
☒ GM Gambia
☒ HR Croatia
☒ HU Hungary
☒ ID Indonesia
☒ IL Israel
☒ IN India
☒ IS Iceland
☒ JP Japan
☒ KE Kenya
☒ KG Kyrgyzstan
☒ KP Democratic People's Republic of Korea
☒ KR Republic of Korea
☒ KZ Kazakhstan
☒ LC Saint Lucia
☒ LK Sri Lanka
☒ LR Liberia
☒ LS Lesotho
☒ LT Lithuania
☒ LU Luxembourg
☒ LV Latvia
☒ MA Morocco
☒ MD Republic of Moldova
☒ MG Madagascar
☒ MK The former Yugoslav Republic of Macedonia
☒ MN Mongolia
☒ MW Malawi
☒ MX Mexico
☒ MZ Mozambique
☒ NO Norway
☒ NZ New Zealand
☒ OM Oman
☒ PH Philippines
☒ PL Poland
☒ PT Portugal
☒ RO Romania
☒ RU Russian Federation
☒ SD Sudan
☒ SE Sweden
☒ SG Singapore
☒ SI Slovenia
☒ SK Slovakia
☒ SL Sierra Leone
☒ TJ Tajikistan
☒ TM Turkmenistan
☒ TN Tunisia
☒ TR Turkey
☒ TT Trinidad and Tobago
☒ TZ United Republic of Tanzania
☒ UA Ukraine
☒ UG Uganda
☒ US United States of America
☒ UZ Uzbekistan
☒ VN Viet Nam
☒ YU Yugoslavia
☒ ZA South Africa
☒ ZM Zambia
☒ ZW Zimbabwe

Check-boxes below reserved for designating States which have become party to the PCT after issuance of this sheet:

☐ ..... ☐ ..... ☐ .....
☐ ..... ☐ ..... ☐ .....

Precautionary Designation Statement: In addition to the designations made above, the applicant also makes under Rule 4.9(b) all other designations which would be permitted under the PCT except any designation(s) indicated in the Supplemental Box as being excluded from the scope of this statement. The applicant declares that those additional designations are subject to confirmation and that any designation which is not confirmed before the expiration of 15 months from the priority date is to be regarded as withdrawn by the applicant at the expiration of that time limit. *(Confirmation (including fees) must reach the receiving Office within the 15-month time limit.)*

See Notes to the request form

Supplemental Box *If the Supplemental Box is not used, this sheet should not be included in the request.*

1. *If, in any of the Boxes, except Boxes Nos. VIII(i) to (v) for which a special continuation box is provided, the space is insufficient to furnish all the information: in such case, write "Continuation of Box No...." (indicate the number of the Box) and furnish the information in the same manner as required according to the captions of the Box in which the space was insufficient, in particular:*

   (i) *if more than two persons are to be indicated as applicants and/or inventors and no "continuation sheet" is available: in such case, write "Continuation of Box No. III" and indicate for each additional person the same type of information as required in Box No. III. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below;*

   (ii) *if, in Box No. II or in any of the sub-boxes of Box No. III, the indication "the States indicated in the Supplemental Box" is checked: in such case, write "Continuation of Box No. II" or "Continuation of Box No. III" or "Continuation of Boxes No. II and No. III" (as the case may be), indicate the name of the applicant(s) involved and, next to (each) such name, the State(s) (and/or, where applicable, ARIPO, Eurasian, European or OAPI patent) for the purposes of which the named person is applicant;*

   (iii) *if, in Box No. II or in any of the sub-boxes of Box No. III, the inventor or the inventor/applicant is not inventor for the purposes of all designated States or for the purposes of the United States of America: in such case, write "Continuation of Box No. II" or "Continuation of Box No. III" or "Continuation of Boxes No. II and No. III" (as the case may be), indicate the name of the inventor(s) and, next to (each) such name, the State(s) (and/or, where applicable, ARIPO, Eurasian, European or OAPI patent) for the purposes of which the named person is inventor;*

   (iv) *if, in addition to the agent(s) indicated in Box No. IV, there are further agents: in such case, write "Continuation of Box No. IV" and indicate for each further agent the same type of information as required in Box No. IV;*

   (v) *if, in Box No. V, the name of any State (or OAPI) is accompanied by the indication "patent of addition," or "certificate of addition," or if, in Box No. V, the name of the United States of America is accompanied by an indication "continuation" or "continuation-in-part": in such case, write "Continuation of Box No. V" and the name of each State involved (or OAPI), and after the name of each such State (or OAPI), the number of the parent title or parent application and the date of grant of the parent title or filing of the parent application;*

   (vi) *if, in Box No. VI, there are more than five earlier applications whose priority is claimed: in such case, write "Continuation of Box No. VI" and indicate for each additional earlier application the same type of information as required in Box No. VI.*

2. *If, with regard to the precautionary designation statement contained in Box No. V, the applicant wishes to exclude any State(s) from the scope of that statement: in such case, write "Designation(s) excluded from precautionary designation statement" and indicate the name or two-letter code of each State so excluded.*

CONTINUATION BOX NO. IV
CARPMAEL, John William Maurice
JONES, Alan John
HOWICK, Nicholas Keith
FISHER, Adrian John
MERCER, Christopher Paul
HALLYBONE, Huw George
JACKSON, Richard Eric
HOWARD, Paul Nicholas
JAMES, Anthony Christopher W.P.
COCKERTON, Bruce Roger

Box No. VI PRIORITY CLAIM

The priority of the following earlier application(s) is hereby claimed:

Filing date of earlier application *(day/month/year)*	Number of earlier application	Where earlier application is:		
		national application: country	regional application:* regional Office	international application: receiving Office
item (1) 11.01.01 11 January 2001	0100750.9	GB		
item (2)				
item (3)				
item (4)				
item (5)				

☐ Further priority claims are indicated in the Supplemental Box.

The receiving Office is requested to prepare and transmit to the International Bureau a certified copy of the earlier application(s) *(only if the earlier application was filed with the Office which for the purposes of this international application is the receiving Office)* identified above as:

☐ all items ☒ item (1) ☐ item (2) ☐ item (3) ☐ item (4) ☐ item (5) ☐ other, see Supplemental Box

* *Where the earlier application is an ARIPO application, indicate at least one country party to the Paris Convention for the Protection of Industrial Property or one Member of the World Trade Organization for which that earlier application was filed (Rule 4.10(b)(ii))*: . . . .

Box No. VII INTERNATIONAL SEARCHING AUTHORITY

Choice of International Searching Authority (ISA) *(if two or more International Searching Authorities are competent to carry out the international search, indicate the Authority chosen; the two-letter code may be used)*:

ISA / EP

Request to use results of earlier search; reference to that search *(if an earlier search has been carried out by or requested from the International Searching Authority)*:

Date *(day/month/year)* Number Country *(or regional Office)*

Box No. VIII DECLARATIONS

The following **declarations** are contained in Boxes Nos. VIII (i) to (v) *(mark the applicable check-boxes below and indicate in the right column the number of each type of declaration)*: Number of declarations

☐ Box No. VIII (i)	Declaration as to the identity of the inventor	:
☐ Box No. VIII (ii)	Declaration as to the applicant's entitlement, as at the international filing date, to apply for and be granted a patent	:
☐ Box No. VIII (iii)	Declaration as to the applicant's entitlement, as at the international filing date, to claim the priority of the earlier application	:
☐ Box No. VIII (iv)	Declaration of inventorship (only for the purposes of the designation of the United States of America)	:
☐ Box No. VIII (v)	Declaration as to non-prejudicial disclosures or exceptions to lack of novelty	:

Box No. IX CHECK LIST; LANGUAGE OF FILING

This international application contains:

(a) the following number of **sheets in paper form**:

request (including declaration sheets)	:	6
description (excluding sequence listing part)	:	71
claims	:	9
abstract	:	1
drawings	:	41
Sub-total number of sheets	:	128
sequence listing part of description *(actual number of sheets if filed in paper form, whether or not also filed in computer readable form; see (b) below)*	:	4
Total number of sheets	:	132

(b) sequence listing part of description filed in **computer readable form**

(i) ☐ only (under Section 801(a)(i))

(ii) ☐ in addition to being filed in paper form (under Section 801(a)(ii))

Type and number of carriers (diskette, CD-ROM, CD-R or other) on which the sequence listing part is contained *(additional copies to be indicated under item 9(ii), in right column)*: .....

This international application is **accompanied by** the following item(s) *(mark the applicable check-boxes below and indicate in right column the number of each item)*:

		Number of items
1. ☒	fee calculation sheet	1
2. ☐	original separate power of attorney	
3. ☐	original general power of attorney	
4. ☒	copy of general power of attorney; reference number, if any:	4
5. ☐	statement explaining lack of signature	
6. ☐	priority document(s) identified in Box No. VI as item(s):	
7. ☐	translation of international application into *(language)*:	
8. ☐	separate indications concerning deposited microorganism or other biological material	
9. ☐	sequence listing in computer readable form (indicate also type and number of carriers (diskette, CD-ROM, CD-R or other))	
	(i) ☐ copy submitted for the purposes of international search under Rule 13*ter* only (and not as part of the international application)	
	(ii) ☐ *(only where check-box (b)(i) or (b)(ii) is marked in left column)* additional copies including, where applicable, the copy for the purposes of international search under Rule 13*ter*	
	(iii) ☐ together with relevant statement as to the identity of the copy or copies with the sequence listing part mentioned in left column	
10. ☒	other *(specify)*: 23/77	1

Figure of the drawings which should accompany the abstract:

Language of filing of the international application: ENGLISH

Box No. X SIGNATURE OF APPLICANT, AGENT OR COMMON REPRESENTATIVE

Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such capacity is not obvious from reading the request).

John Wm Carpmael

MERCER, Christopher Paul

Authorised Representative

For receiving Office use only

1. Date of actual receipt of the purported international application:	11 JANUARY 2002 11/01/02	2. Drawings: ☑ received:
3. Corrected date of actual receipt due to later but timely received papers or drawings completing the purported international application:		
4. Date of timely receipt of the required corrections under PCT Article 11(2):		☐ not received:
5. International Searching Authority (if two or more are competent): ISA /	6. ☐ Transmittal of search copy delayed until search fee is paid	

For International Bureau use only

Date of receipt of the record copy by the International Bureau:

ADHESION MOLECULES

This invention relates to novel proteins, termed AAC74854.1, AAC76768.1 and P10155 herein identified as adhesion molecules and to the use of these proteins and nucleic acid sequences from the encoding genes in the diagnosis, prevention and treatment of disease.

All publications, patents and patent applications cited herein are incorporated in full by reference.

BACKGROUND

The process of drug discovery is presently undergoing a fundamental revolution as the era of functional genomics comes of age. The term "functional genomics" applies to an approach utilising bioinformatics tools to ascribe function to protein sequences of interest. Such tools are becoming increasingly necessary as the speed of generation of sequence data is rapidly outpacing the ability of research laboratories to assign functions to these protein sequences.

As bioinformatics tools increase in potency and in accuracy, these tools are rapidly replacing the conventional techniques of biochemical characterisation. Indeed, the advanced bioinformatics tools used in identifying the present invention are now capable of outputting results in which a high degree of confidence can be placed.

Various institutions and commercial organisations are examining sequence data as they become available and significant discoveries are being made on an on-going basis. However, there remains a continuing need to identify and characterise further genes and the polypeptides that they encode, as targets for research and for drug discovery.

Recently, a remarkable tool for the evaluation of sequences of unknown function has been developed by the Applicant for the present invention. This tool is a database system, termed the Biopendium search database, that is the subject of co-pending International Patent Application No. PCT/GB01/01105. This database system consists of an integrated data resource created using proprietary technology and containing information generated from an all-by-all comparison of all available protein or nucleic acid sequences.

The aim behind the integration of these sequence data from separate data resources is to combine as much data as possible, relating both to the sequences themselves and to

information relevant to each sequence, into one integrated resource. All the available data relating to each sequence, including data on the three-dimensional structure of the encoded protein, if this is available, are integrated together to make best use of the information that is known about each sequence and thus to allow the most educated predictions to be made from comparisons of these sequences. The annotation that is generated in the database and which accompanies each sequence entry imparts a biologically relevant context to the sequence information.

This data resource has made possible the accurate prediction of protein function from sequence alone. Using conventional technology, this is only possible for proteins that exhibit a high degree of sequence identity (above about 20%-30% identity) to other proteins in the same functional family. Accurate predictions are not possible for proteins that exhibit a very low degree of sequence homology to other related proteins of known function.

In the present case, a protein whose sequence is recorded in a publicly available database as AAC74854.1 (NCBI Genebank nucleotide accession number AE000273 and a Genebank protein accession number AAC74854.1), is implicated as a novel member of the adhesion molecule family.

A second protein whose sequence is recorded in a publicly available database as AAC76768.1 (NCBI Genebank nucleotide accession number AE000451 and a Genebank protein accession number AAC76768.1), is also implicated as a novel member of the adhesion molecule family.

A third protein whose sequence is recorded in a publicly available database as P10155 (NCBI Genebank nucleotide accession number J04137 and a Genebank protein accession number 10155), is also implicated as a novel member of the adhesion molecule family.

Introduction to adhesion molecules

Adhesion molecules are involved in a range of biological processes, including embryogenesis (Martin-Bermudo, M.D., *et al.,*. Development. 2000 Jun;127(12): 2607-15; Chen, L.M., *et al.*, J Neurosci. 2000 May 15;20(10): 3776-84; Zweegman, S., *et al.*, Exp Hematol. 2000 Apr;28(4): 401-10; Darribere, T., *et al.*, Biol Cell. 2000 Jan;92(1): 5-25),

maintenance of tissue integrity (Eckes, B., *et al.,* J Cell Sci. 2000;113(Pt 13): 2455-2462; Buckwalter, J.A., *et al.,* Instr Course Lect. 2000;49: 481-9; Frenette, P.S., *et al.,*. J Exp Med. 2000 Apr 17;191(8): 1413-22; Delmas, V., *et al.,* Dev Biol. 1999 Dec 15;216(2): 491-506; Humphries, M.J., *et al.,* Trends Pharmacol Sci. 2000 Jan;21(1): 29-32; Miosge, N., *et al.,* Lab Invest. 1999 Dec;79(12): 1591-9), leukocyte extravasation/inflammation (Lim, L.H., *et al.* Am J Respir Cell Mol Biol. 2000 Jun;22(6): 693-701; Johnston, B., *et al.,* Microcirculation. 2000 Apr;7(2): 109-18; Mertens, A.V., *et al.,* Clin Exp Allergy. 1993 Oct;23(10): 868-73; Chcialowski, A., *et al.,* Pol Merkuriusz Lek. 2000 Jan;7(43): 13-7; Rojas, A.I., *et al.,* Crit Rev Oral Biol Med. 1999;10(3): 337-58; Marinova-Mutafchieva, L., *et al.,* Arthritis Rheum. 2000 Mar;43(3): 638-44; Vijayan, K.V., *et al.,* J Clin Invest. 2000 Mar;105(6): 793-802; Currie, A.J., *et al,*. J Immunol. 2000 Apr 1;164(7): 3878-86; Rowin, M.E., et al., Inflammation. 2000 Apr;24(2): 157-73; Johnston, B., *et al.,* J Immunol. 2000 Mar 15;164(6): 3337-44; Gerst, J.L., *et al.,* J Neurosci Res. 2000 Mar 1;59(5): 680-4; Kagawa, T.F., *et al.,* Proc Natl Acad Sci U S A. 2000 Feb 29;97(5): 2235-40; Hillan, K.J., *et al.,* Liver. 1999 Dec;19(6): 509-18; Panes, J., 1999 Dec;22(10): 514-24; Arao, T., *et al.,* J Clin Endocrinol Metab. 2000 Jan;85(1): 382-9; Souza, H.S., *et al.,* Gut. 1999 Dec;45(6): 856-63; Grunstein, M.M., *et al.,* Am J Physiol Lung Cell Mol Physiol. 2000 Jun;278(6): L1154-63; Mertens, A.V., *et al.,* Clin Exp Allergy. 1993 Oct;23(10): 868-73; Berends, C., *et al.,* Clin Exp Allergy. 1993 Nov;23(11): 926-33; Fernvik, E., *et al.,* Inflammation. 2000 Feb;24(1): 73-87; Bocchino, V., *et al.,* J Allergy Clin Immunol. 2000 Jan;105(1 Pt 1): 65-70), oncogenesis (Orr, F.W., *et al.,* Cancer. 2000 Jun;88(S12): 2912-2918; Zeller, W., *et al.,* J Hematother Stem Cell Res. 1999 Oct;8(5): 539-46; Okada, T., *et al.,* Clin Exp Metastasis. 1999;17(7): 623-9; Mateo, V., *et al.,* Nat Med. 1999 Nov;5(11): 1277-84; Yamaguchi, K., *et al.,* J Exp Clin Cancer Res. 2000 Mar;19(1): 113-20; Maeshima, Y., *et al.,* J Biol Chem. 2000 Jun 2 (epub ahead of reprint); Van Waes, C., *et al.,* Int J Oncol. 2000 Jun;16(6): 1189-95; Damiano, J.S., *et al.*, Leuk Lymphoma. 2000 Jun;38(1-2): 71-81; Seftor, R.E.,*et al.,*Cancer Metastasis Rev. 1999;18(3): 359-75; Shaw, L.M., J Mammary Gland Biol Neoplasia. 1999 Oct;4(4): 367-76; Weyant, M.J., *et al.,* Clin Cancer Res. 2000 Mar;6(3): 949-56), and thrombogenesis (Wang, Y.G., *et al.,* J Physiol (Lond). 2000 Jul 1;526(Pt 1): 57-68; Matsuno, H., *et al.,* Nippon Yakurigaku Zasshi. 2000 Mar;115(3): 143-50; Eliceiri, B.P., *et al.,* Cancer J Sci Am. 2000 May;6 Suppl 3: S245-9; von Beckerath, N., *et al.,* Blood.

2000 Jun 1;95(11): 3297-301; Topol, E.J., *et al.,* Am Heart J. 2000 Jun;139(6): 927-33; Kroll, H., *et al.,* Thromb Haemost. 2000 Mar;83(3): 392-6).

The detailed characterisation of the structure and function of several adhesion-receptor families has led to active programs by a number of pharmaceutical companies to develop adhesion molecule antagonists for use in the treatment of inflammation, oncology, and cardiovascular disease. Adhesion receptors are involved in virtually every aspect of biology from embryogenesis to apoptosis. They are essential to the structural integrity and homeostatic functioning of most tissues. It is therefore not surprising that defects in adhesion receptors cause disease and that many diseases involve modulation of adhesion molecule function.

The Adhesion molecule family in fact represents at least four distinct families which are unified by their function rather than their structure. Of the four families, three are of pharmaceutical interest due to small molecule tractability. They are:

1.The integrin family is a superfamily of α and β heterodimeric transmembrane glycoproteins and is the family, which has attracted most pharmaceutical interest. Its members are large, heavily glycosylated, heterodimeric proteins composed of one of at least 15 distinct α-subunits in non-covalent linkage with one of at least 8 β-subunits. Adhesion receptors bind ligands expressed on cell surfaces, extracellular matrix molecules, and soluble molecules. Integrins are subcategorised based on their β-subunit usage. The members of this family are summarised below in Table 1.

2.Selectins are a small family of three members P, E and L selectin. They are glycoproteins, selectively expressed on cells related to the vasculature, and contain a lectin-binding domain. The members of this family are described below in Table 2.

3.The immunoglobulin family represents the counter receptor for the integrins and includes the intracellular adhesion molecules (ICAMs) and vascular cell adhesion molecules (VCAMs). Members are composed of variable numbers of globular, immunoglobulin-like, extracellular domains. Some members of the family, for example, PECAM-1 (CD31) and NCAM, mediate homotypic adhesion. Some members of the family, for

example ICAM-1 and VCAM-1, mediate adhesion via interactions with integrins. The members of this family are described below in Table 3.

Adhesion molecules have been shown to play a role in diverse physiological functions, many of which can play a role in disease processes. Alteration of their activity is a means to alter the disease phenotype and as such identification of novel adhesion molecules is highly relevant as they may play a role in many diseases, particularly inflammatory disease, oncology, and cardiovascular disease.

Table 1. ***Integrins:***

Integrin Receptor	*Ligand*	*Distribution*
β1 (CD29)		
α1β1	Laminin, Collagen	Activated T cells, fibroblasts
α2β1	Collagen, Laminin	Activated T cells, endothelial cells, platelets, basophils.
α3β1	Laminin, Collagen, Fibronectin	Basement membrane
α4β1	VCAM-1 (domains 1 and 4), Fibronectin (CS-1 domain), MadCAM-1	Lymphocytes, monocytes, eosinophils, basophils, mast cells, NK cells
α5β1	Fibronectin	Lymphocytes, monocytes, endothelial cells, basophils, mast cells, fibroblasts
α6β1	Laminin	Platelets, T cells, eosinophils, monocytes, endothelial cells
α9β1	Tenascin, VCAM-1, Osteopontin	Airway epithelial cells, smooth muscle cells, neutrophils
αVβ1	Vitronectin, fibronectin	Platelets, B cells.
β2 (CD18)		
LFA-1 (CD11a/CD18)	ICAM-1, 2, 3	All leukocytes
Mac-1 (CD11b/CD18)	ICAM-1, Fibrinogen, LPS	Granulocytes, monocytes
αD	ICAM-3, VCAM-1	Tissue macrophages, monocytes, CD8+ T cells, eosinophils

Integrin Receptor	*Ligand*	*Distribution*
β3 (CD61)		
GpIIb/IIIa	Fibrinogen, Vitronectin, Fibronectin, vWF	Platelets, endothelial cells
αV/IIIa	Vitronectin, Fibrinogen, vWF, Laminin, Thrombospondin, Osteopontin	Platelets,
β7		
α4β7 (LPAM-1)	MAdCAM-1, VCAM-1, Firbonectin (CS-1 domain)	Subset of memory T cells, eosinophils, basophils, endothelial cells
αEβ7	E-cadherin	Intestinal intraepithelial lymphocytes.

Table 2. ***Selectins:***

Receptor	*Ligand*	*Distribution*
E-selectin	Sialyl-LewisX, selectin, LFA-1, ESL-1, PSGL-1	L-Activated endothelial cells
L-selectin	GlyCAM-1, MAdCAM-1, CD34, Sialyl LewisX, E-selectin, P-selectin	Resting leukocytes
P-selectin	Sialyl-LewisX, selectin, PSGL-1	L-Activated endothelial cells, activated platelets

Table 3. ***Immunoglobulin superfamily:***

Receptor	*Ligand*	*Distribution*
ICAM-1 5 Ig domains	LFA-1 (CD11a/CD18) Mac-1 (CD11b/CD18), CD43	Widespread, endothelial cells, fibroblasts, epithelium, monocytes, lymphocytes, dendritic cells, chondrocytes.
ICAM-2 2 Ig domains	LFA-1 (CD11b)	endothelial cells (high): lymphocytes, monocytes, basophils, platelets (low).
ICAM-3 5 Ig domains	LFA-1 (αd/CD18)	Lymphocytes, monocytes, neutrophils, eosinophils, basophils.
VCAM-1 6 or 7 Ig domains	α4β1, α4β7	Endothelial cells, monocytes, fibroblasts, dendritic cells, bone marrow stromal cells, myoblasts.
LFA-3 6 Ig domains	CD2	Endothelial cells, leukocytes, epithelial cells
PECAM-1	CD31, heparin	Endothelial cells (at EC-EC junctions), T cell

(CD31)		subsets, platelets, neutrophils, eosinophils, monocytes, smooth muscle cells, bone marrow stem cells.
NCAM	NCAM, heparin SO_4	Neural cells, muscle
MadCAM-1 4 Ig domains	$\alpha 4\beta 7$, L-selectin	Peyer's patch, mesenteric lymph nodes, mucosal endothelial cells, spleen.
CD2	CD58, CD59, CD48	T lymphocytes

THE INVENTION

The invention is based on the discovery that the AAC74854.1 protein, AAC76768.1 protein and P10155 protein function as adhesion molecules.

For the AAC74854.1 protein, it has been found that a region including residues 250-365 of this protein sequence adopts an equivalent fold to residues 132 to 226 of the 1LFA:A (PDB code 1LFA). 1LFA:A is known to function as an adhesion molecule. Furthermore, the divalent metal ion binding residues SER139, SER141 and ASP239 of the 1LFA:A are conserved as SER258, SER260 and ASP348 in AAC74854.1, respectively. This relationship is not just to the 1LFA:A structure, but rather to the adhesion molecule family as a whole. It has been found that a region whose boundaries extend between residue 250 and residue 365 of AAC74854.1 adopts an equivalent fold to to a range of other adhesion molecules including 1BHO. Furthermore, the divalent metal ion binding residues of 1BHO are conserved as SER258, SER260 and ASP348 in AAC74854.1, respectively. Thus, by reference to the Genome Threader^{TM} alignment of AAC74854.1 with the 1LFA:A (1LFA) SER258, SER260 and ASP348 of AAC74854.1 are predicted to form the divalent metal ion binding residues.

The combination of equivalent fold and conservation of divalent metal ion binding residues allows the functional annotation of this region of AAC74854.1, and therefore proteins that include this region, as possessing adhesion molecule activity.

For the AAC76768.1 protein, it has been found that a region including residues 267-384 of this protein sequence adopts an equivalent fold to residues 147 to 284 of the 1AOX:A (PDB code 1AOX). 1AOX:A is known to function as an adhesion molecule.

Furthermore, the divalent metal ion binding residues SER153, SER155 and ASP254 of the 1AOX:A are conserved as SER273, SER275 and ASP365 in AAC76768.1, respectively. This relationship is not just to 1AOX:A, but rather to the adhesion molecule family as a whole. It has been found that a region whose boundaries extend between residue 267 and residue 384 of AAC76768.1 adopts an equivalent fold to to a range of other adhesion molecules including the 1BHO structure. Furthermore, divalent metal ion binding residues of 1BHO are conserved as SER273, SER275 and ASP365 in AAC76768.1, respectively. Thus, by reference to the Genome Threader™ alignment of AAC76768.1 with the 1AOX:A (1AOX) SER273, SER275 and ASP365 of AAC76768.1 are predicted to form the divalent metal ion binding residues.

The combination of equivalent fold and conservation of the divalent metal ion binding residues allows the functional annotation of this region of AAC76768.1, and therefore proteins that include this region, as possessing adhesion molecule activity.

For the P10155 protein, it has been found that a region including residues 373-503 of this protein sequence adopts an equivalent fold to residues 137 to 273 of the 1JLM (PDB code 1JLM). 1JLM is known to function as an adhesion molecule. Furthermore, the divalent metal ion binding residues SER142, SER144 and ASP242 of the 1JLM are conserved as SER378, SER380 and ASP469 in P10155, respectively. This relationship is not just to 1JLM, but rather to the adhesion molecule family as a whole. It has been found that a region whose boundaries extend between residue 373 and residue 503 of P10155 adopts an equivalent fold to to a range of other adhesion molecules including the 1BHO structure. Furthermore, divalent metal ion binding residues of 1BHO are conserved as SER378, SER380 and ASP469 in P10155, respectively. Thus, by reference to the Genome Threader™ alignment of P10155 with the 1JLM (1JLM) SER378, SER380 and ASP469 of P10155 are predicted to form thedivalent metal ion binding residues.

The combination of equivalent fold and conservation of divalent metal ion binding residues allows the functional annotation of this region of P10155, and therefore proteins that include this region, as possessing adhesion molecule activity.

In a first aspect, the invention provides a polypeptide, which polypeptide:

(i) comprises the amino acid sequence as recited in SEQ ID NO:2, SEQ ID NO:4, or

SEQ ID NO:6;

(ii) is a fragment thereof having adhesion molecule activity or having an antigenic determinant in common with the polypeptides of (i); or

(iii) is a functional equivalent of (i) or (ii).

Preferably, a polypeptide acording to the present invention consists of the amino acid sequence as recited in SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6.

The polypeptide having the sequence recited in SEQ ID NO:2 is referred to hereafter as "the ADS1 polypeptide".

According to this aspect of the invention, a preferred polypeptide fragment according to part ii) above includes the region of the ADS1 polypeptide that is predicted as that responsible for adhesion molecule activity (hereafter, the "ADS1 adhesion molecule region"), or is a variant thereof that possesses the divalent metal ion binding (SER258, SER260 and ASP348, or equivalent residues). As defined herein, the ADS1 adhesion molecule region is considered to extend between residue 250 and residue 365 of the ADS1 polypeptide sequence.

The polypeptide having the sequence recited in SEQ ID NO:4 is referred to hereafter as "the ADS2 polypeptide".

According to this aspect of the invention, a preferred polypeptide fragment according to part ii) above includes the region of the ADS2 polypeptide that is predicted as that responsible for adhesion molecule activity (hereafter, the "ADS2 adhesion molecule region"), or is a variant thereof that possesses the divalent metal ion binding (SER273, SER275 and ASP365, or equivalent residues). As defined herein, the ADS2 adhesion molecule region is considered to extend between residue 267 and residue 384 of the ADS2 polypeptide sequence.

The polypeptide having the sequence recited in SEQ ID NO:6 is referred to hereafter as "the ADS5 polypeptide".

According to this aspect of the invention, a preferred polypeptide fragment according to part ii) above includes the region of the ADS5 polypeptide that is predicted as that responsible for adhesion molecule activity (hereafter, the "ADS5 adhesion molecule

region"), or is a variant thereof that possesses the divalent metal ion binding (SER378, SER380 and ASP365, or equivalent residues). As defined herein, the ADS5 adhesion molecule region is considered to extend between residue 373 and residue 503 of the ADS5 polypeptide sequence.

This aspect of the invention also includes fusion proteins that incorporate polypeptide fragments and variants of these polypeptide fragments as defined above, provided that said fusion proteins possess activity as an adhesion molecule.

In a second aspect, the invention provides a purified nucleic acid molecule that encodes a polypeptide of the first aspect of the invention. Preferably, the purified nucleic acid molecule has the nucleic acid sequence as recited in SEQ ID NO:1 (encoding the ADS1 polypeptide), SEQ ID NO:3 (encoding the ADS2 polypeptide), or SEQ ID NO:5 (encoding the ADS5 polypeptide), or is a redundant equivalent or fragment of any one of these sequences. A preferred nucleic acid fragment is one that encodes a polypeptide fragment according to part ii) above, preferably a polypeptide fragment that includes the ADS1 adhesion molecule region, the ADS2 adhesion molecule region, the ADS5 adhesion molecule region, or that encodes a variant of these fragments as this term is defined above.

In a third aspect, the invention provides a purified nucleic acid molecule that hybridizes under high stringency conditions with a nucleic acid molecule of the second aspect of the invention.

In a fourth aspect, the invention provides a vector, such as an expression vector, that contains a nucleic acid molecule of the second or third aspect of the invention.

In a fifth aspect, the invention provides a host cell transformed with a vector of the fourth aspect of the invention.

In a sixth aspect, the invention provides a ligand which binds specifically to, and which preferably inhibits the adhesion molecule activity of a polypeptide of the first aspect of the invention.

In a seventh aspect, the invention provides a compound that is effective to alter the expression of a natural gene which encodes a polypeptide of the first aspect of the

invention or to regulate the activity of a polypeptide of the first aspect of the invention.

A compound of the seventh aspect of the invention may either increase (agonise) or decrease (antagonise) the level of expression of the gene or the activity of the polypeptide. Importantly, the identification of the function of the region defined herein as the ADS1, ADS2 and ADS5 adhesion molecule regions of the ADS1, ADS2 and ADS5 polypeptides, respectively, allows for the design of screening methods capable of identifying compounds that are effective in the treatment and/or diagnosis of diseases in which adhesion molecules are implicated.

In an eighth aspect, the invention provides a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, or a composition of the eleventh aspect of the invention, for use in therapy or diagnosis. These molecules may also be used in the manufacture of a medicament for the treatment of cardiovascular diseases including atherosclerosis, ischaemia, restenosis, reperfusion injury, sepsis, haematological diseases such as leukaemia, blood clotting disorders, such as thrombosis, cancer including lung, prostate, breast, colorectal and brain tumours, metastasis, inflammatory diseases such as rhinitis, gastrointestinal diseases, including inflammatory bowel disease, ulcerative colitis, Crohn's disease, respiratory diseases including asthma, chronic obstructive pulmonary disease (COPD), respiratory distress syndrome, pulmonary fibrosis, immune disorders, including autoimmune diseases, rheumatoid arthritis, transplant rejection, allergy, liver diseases such as cirrhosis, endocrine diseases, such as diabetes, bone diseases such as osteoporosis, neurological diseases including stroke, multiple sclerosis, spinal cord injury, burns and wound healing, infections, preferably bacterial infection and most preferably *E. coli* infection.

In a ninth aspect, the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide of the first aspect of the invention or the activity of a polypeptide of the first aspect of the invention in tissue from said patient and comparing said level of expression or activity to a control level, wherein a level that is different to said control level is indicative of

disease. Such a method will preferably be carried out *in vitro*. Similar methods may be used for monitoring the therapeutic treatment of disease in a patient, wherein altering the level of expression or activity of a polypeptide or nucleic acid molecule over the period of time towards a control level is indicative of regression of disease.

The adhesion molecule whose sequences are presented in SEQ ID NO: 2 and SEQ ID NO: 4 are implicated herein in the pathogenicity of the organism *Escherichia Coli*. Accordingly, ligands to this polypeptide, and in particular, to the adhesion molecule regions of the ADS1 and ADS2 polypeptides respectively, as these regions are defined herein, are likely to be effective in controlling disease caused by this organism. Furthermore, these polypeptides, and in particular, polypeptide fragments including the adhesion molecule regions of the ADS1 and ADS2 polypeptide sequences provide a potential component for a vaccine against this organism and the diseases that it causes.

A preferred method for detecting polypeptides of the first aspect of the invention comprises the steps of: (a) contacting a ligand, such as an antibody, of the sixth aspect of the invention with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

A number of different such methods according to the ninth aspect of the invention exist, as the skilled reader will be aware, such as methods of nucleic acid hybridization with short probes, point mutation analysis, polymerase chain reaction (PCR) amplification and methods using antibodies to detect aberrant protein levels. Similar methods may be used on a short or long term basis to allow therapeutic treatment of a disease to be monitored in a patient. The invention also provides kits that are useful in these methods for diagnosing disease.

In a tenth aspect, the invention provides for the use of a polypeptide of the first aspect of the invention as an adhesion molecule. The invention also provides for the use of a nucleic acid molecule according to the second or third aspects of the invention to express a protein that possesses adhesion molecule activity. The invention also provides a method for effecting adhesion molecule activity, said method utilising a polypeptide of the first aspect of the invention.

In an eleventh aspect, the invention provides a pharmaceutical composition comprising a

polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, in conjunction with a pharmaceutically-acceptable carrier.

In a twelfth aspect, the present invention provides a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, or a composition of the eleventh aspect of the invention, for use in the manufacture of a medicament for the diagnosis or treatment of a disease, such as cardiovascular diseases including atherosclerosis, ischaemia, restenosis, reperfusion injury, sepsis, haematological diseases such as leukaemia, blood clotting disorders, such as thrombosis, cancer including lung, prostate, breast, colorectal and brain tumours, metastasis, inflammatory diseases such as rhinitis, gastrointestinal diseases, including inflammatory bowel disease, ulcerative colitis, Crohn's disease, respiratory diseases including asthma, chronic obstructive pulmonary disease (COPD), respiratory distress syndrome, pulmonary fibrosis, immune disorders, including autoimmune diseases, rheumatoid arthritis, transplant rejection, allergy, liver diseases such as cirrhosis, endocrine diseases, such as diabetes, bone diseases such as osteoporosis, neurological diseases including stroke, multiple sclerosis, spinal cord injury, burns and wound healing, infections, preferably bacterial infection and most preferably E coli infection.

In a thirteenth aspect, the invention provides a method of treating a disease in a patient comprising administering to the patient a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention.

For diseases in which the expression of a natural gene encoding a polypeptide of the first aspect of the invention, or in which the activity of a polypeptide of the first aspect of the invention, is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand or compound

administered to the patient should be an agonist. Conversely, for diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand or compound administered to the patient should be an antagonist. Examples of such antagonists include antisense nucleic acid molecules, ribozymes and ligands, such as antibodies.

In a fourteenth aspect, the invention provides transgenic or knockout non-human animals that have been transformed to express higher, lower or absent levels of a polypeptide of the first aspect of the invention. Such transgenic animals are very useful models for the study of disease and may also be using in screening regimes for the identification of compounds that are effective in the treatment or diagnosis of such a disease.

A summary of standard techniques and procedures which may be employed in order to utilise the invention is given below. It will be understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors and reagents described. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and it is not intended that this terminology should limit the scope of the present invention. The extent of the invention is limited only by the terms of the appended claims.

Standard abbreviations for nucleotides and amino acids are used in this specification.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA technology and immunology, which are within the skill of the those working in the art.

Such techniques are explained fully in the literature. Examples of particularly suitable texts for consultation include the following: Sambrook Molecular Cloning; A Laboratory Manual, Second Edition (1989); DNA Cloning, Volumes I and II (D.N. Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed. 1984); Nucleic Acid Hybridization (B.D. Hames & S.J. Higgins eds. 1984); Transcription and Translation (B.D. Hames & S.J. Higgins eds. 1984); Animal Cell Culture (R.I. Freshney ed. 1986); Immobilized Cells and Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide to Molecular Cloning (1984); the Methods in Enzymology series (Academic Press, Inc.), especially volumes 154 &

155; Gene Transfer Vectors for Mammalian Cells (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Immunochemical Methods in Cell and Molecular Biology (Mayer and Walker, eds. 1987, Academic Press, London); Scopes, (1987) Protein Purification: Principles and Practice, Second Edition (Springer Verlag, N.Y.); and Handbook of Experimental Immunology, Volumes I-IV (D.M. Weir and C. C. Blackwell eds. 1986).

As used herein, the term "polypeptide" includes any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e. peptide isosteres. This term refers both to short chains (peptides and oligopeptides) and to longer chains (proteins).

The polypeptide of the present invention may be in the form of a mature protein or may be a pre-, pro- or prepro- protein that can be activated by cleavage of the pre-, pro- or prepro- portion to produce an active mature polypeptide. In such polypeptides, the pre-, pro- or prepro- sequence may be a leader or secretory sequence or may be a sequence that is employed for purification of the mature polypeptide sequence.

The polypeptide of the first aspect of the invention may form part of a fusion protein. For example, it is often advantageous to include one or more additional amino acid sequences which may contain secretory or leader sequences, pro-sequences, sequences which aid in purification, or sequences that confer higher protein stability, for example during recombinant production. Alternatively or additionally, the mature polypeptide may be fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol).

Polypeptides may contain amino acids other than the 20 gene-encoded amino acids, modified either by natural processes, such as by post-translational processing or by chemical modification techniques which are well known in the art. Among the known modifications which may commonly be present in polypeptides of the present invention are glycosylation, lipid attachment, sulphation, gamma-carboxylation, for instance of glutamic acid residues, hydroxylation and ADP-ribosylation. Other potential modifications include acetylation, acylation, amidation, covalent attachment of flavin, covalent attachment of a haeme moiety, covalent attachment of a nucleotide or nucleotide

derivative, covalent attachment of a lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulphide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, GPI anchor formation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl terminus in a polypeptide, or both, by a covalent modification is common in naturally-occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention.

The modifications that occur in a polypeptide often will be a function of how the polypeptide is made. For polypeptides that are made recombinantly, the nature and extent of the modifications in large part will be determined by the post-translational modification capacity of the particular host cell and the modification signals that are present in the amino acid sequence of the polypeptide in question. For instance, glycosylation patterns vary between different types of host cell.

The polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally-occurring polypeptides (for example purified from cell culture), recombinantly-produced polypeptides (including fusion proteins), synthetically-produced polypeptides or polypeptides that are produced by a combination of these methods.

The functionally-equivalent polypeptides of the first aspect of the invention may be polypeptides that are homologous to the ADS1, ADS2, or ADS5 polypeptides or to the the adhesion molecule regions of the ADS1, ADS2, or ADS5 polypeptides. Two polypeptides are said to be "homologous", as the term is used herein, if the sequence of one of the polypeptides has a high enough degree of identity or similarity to the sequence of the other polypeptide. "Identity" indicates that at any particular position in the aligned sequences, the amino acid residue is identical between the sequences. "Similarity"

indicates that, at any particular position in the aligned sequences, the amino acid residue is of a similar type between the sequences. Degrees of identity and similarity can be readily calculated (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing. Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part 1, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991).

Homologous polypeptides therefore include natural biological variants (for example, allelic variants or geographical variations within the species from which the polypeptides are derived) and mutants (such as mutants containing amino acid substitutions, insertions or deletions) of the ADS1, ADS2, or ADS5 polypeptides. Such mutants may include polypeptides in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; among the basic residues Lys and Arg; or among the aromatic residues Phe and Tyr. Particularly preferred are variants in which several, i.e. between 5 and 10, 1 and 5, 1 and 3, 1 and 2 or just 1 amino acids are substituted, deleted or added in any combination. Especially preferred are silent substitutions, additions and deletions, which do not alter the properties and activities of the protein. Also especially preferred in this regard are conservative substitutions.

Such mutants also include polypeptides in which one or more of the amino acid residues includes a substituent group.

Typically, greater than 30% identity between two polypeptides (preferably, over a specified region) is considered to be an indication of functional equivalence. Preferably, functionally equivalent polypeptides of the first aspect of the invention have a degree of sequence identity with the ADS1, ADS2, or ADS5 polypeptide, or the adhesion molecule

regsion thereof, or with active fragments thereof, of greater than 30%. More preferred polypeptides have degrees of identity of greater than 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98% or 99%, respectively with the ADS1, ADS2, or ADS5 polypeptide, or the adhesion molecule regsion thereof, or with active fragments thereof.

Percentage identity, as referred to herein, is as determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for Biotechnology Information; http://www.ncbi.nlm.nih.gov/) [Blosum 62 matrix; gap open penalty=11 and gap extension penalty=1].

In the present case, preferred active fragments of the ADS1 polypeptide are those that include the ADS1 adhesion molecule region and which possess the divalent metal ion binding of residues SER258, SER260 and ASP348, or equivalent residues. By "equivalent residues" is meant residues that are equivalent to the divalent metal ion binding residues, provided that the adhesion molecule region retains activity as an adhesion molecule. For example serine may be replaced by threonine, or aspartate may be replaced by glutamate. Accordingly, this aspect of the invention includes polypeptides that have degrees of identity of greater than 30%, preferably, greater than 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98% or 99%, respectively, with the adhesion molecule region of the ADS1 polypeptide and which possess the divalent metal ion binding of SER258, SER260 and ASP348, or equivalent residues. As discussed above, the ADS1 adhesion molecule region is considered to extend between residue 250 and residue 365 of the ADS1 polypeptide sequence.

In the present case, preferred active fragments of the ADS2 polypeptide are those that include the ADS2 adhesion molecule region and which possess the divalent metal ion binding of residues SER273, SER275 and ASP365, or equivalent residues. By "equivalent residues" is meant residues that are equivalent to the divalent metal ion binding residues, provided that the adhesion molecule region retains activity as an adhesion molecule. For example serine may be replaced by threonine, or aspartate may be replaced by glutamate. Accordingly, this aspect of the invention includes polypeptides that have degrees of identity of greater than 30%, preferably, greater than 40%, 50%, 60%, 70%, 80%,90%, 95%, 98% or 99%, respectively, with the adhesion molecule

region of the ADS2 polypeptide and which possess the divalent metal ion binding of SER273, SER275 and ASP365, or equivalent residues. As discussed above, the ADS2 adhesion molecule region is considered to extend between residue 267 and residue 384 of the ADS2 polypeptide sequence.

In the present case, preferred active fragments of the ADS5 polypeptide are those that include the ADS5 adhesion molecule region and which possess the divalent metal ion binding of residues SER378, SER380 and ASP469, or equivalent residues. By "equivalent residues" is meant residues that are equivalent to the divalent metal ion binding residues, provided that the adhesion molecule region retains activity as an adhesion molecule. For example serine may be replaced by threonine, or aspartate may be replaced by glutamate. Accordingly, this aspect of the invention includes polypeptides that have degrees of identity of greater than 30%, preferably, greater than 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98% or 99%, respectively, with the adhesion molecule region of the ADS5 polypeptide and which possess the divalent metal ion binding of SER378, SER380 and ASP469, or equivalent residues. As discussed above, the ADS5 adhesion molecule region is considered to extend between residue 373 and residue 503 of the ADS5 polypeptide sequence.

The functionally-equivalent polypeptides of the first aspect of the invention may also be polypeptides which have been identified using one or more techniques of structural alignment. For example, the Inpharmatica Genome Threader™ technology that forms one aspect of the search tools used to generate the Biopendium search database may be used (see co-pending International patent application PCT/GB01/01105) to identify polypeptides of presently-unknown function which, while having low sequence identity as compared to the ADS1, ADS2 or ADS5 polypeptides, are predicted to have adhesion molecule activity, by virtue of sharing significant structural homology with the ADS1, ADS2 or ADS5 polypeptide sequences.

By "significant structural homology" is meant that the Inpharmatica Genome Threader™ predicts two proteins, or protein regions, to share structural homology with a certainty of at least 10% more preferably, at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and above. The certainty value of the Inpharmatica Genome Threader™ is calculated as

follows. A set of comparisons was initially performed using the Inpharmatica Genome Threader™ exclusively using sequences of known structure. Some of the comparisons were between proteins that were known to be related (on the basis of structure). A neural network was then trained on the basis that it needed to best distinguish between the known relationships and known not-relationships taken from the CATH structure classification (www.biochem.ucl.ac.uk/bsm/cath). This resulted in a neural network score between 0 and 1. However, again as the number of proteins that are related and the number that are unrelated were known, it was possible to partition the neural network results into packets and calculate empirically the percentage of the results that were correct. In this manner, any genuine prediction in the Biopendium search database has an attached neural network score and the percentage confidence is a reflection of how successful the Inpharmatica Genome Threader™ was in the training/testing set.

Structural homologues of ADS1 should share structural homology with the ADS1 adhesion molecule region and possess the divalent metal ion binding residues SER258, SER260 and ASP348, or equivalent residues. Such structural homologues are predicted to have adhesion molecule activity by virtue of sharing significant structural homology with this polypeptide sequence and possessing the divalent metal ion binding residues.

Structural homologues of ADS2 should share structural homology with the ADS2 adhesion molecule region and possess the divalent metal ion binding residues SER273, SER275 and ASP365, or equivalent residues. Such structural homologues are predicted to have adhesion molecule activity by virtue of sharing significant structural homology with this polypeptide sequence and possessing the divalent metal ion binding residues.

Structural homologues of ADS5 should share structural homology with the ADS5 adhesion molecule region and possess the divalent metal ion binding residues SER378, SER380 and ASP469, or equivalent residues. Such structural homologues are predicted to have adhesion molecule activity by virtue of sharing significant structural homology with this polypeptide sequence and possessing the divalent metal ion binding residues.

The polypeptides of the first aspect of the invention also include fragments of the ADS1, ADS2, and ADS5 polypeptides, functional equivalents of the fragments of the ADS1, ADS2, and ADS5 polypeptides, and fragments of the functional equivalents of the ADS1,

ADS2, and ADS5 polypeptides, provided that those functional equivalents and fragments retain adhesion molecule activity or have an antigenic determinant in common with the ADS1, ADS2, or ADS5 polypeptides.

As used herein, the term "fragment" refers to a polypeptide having an amino acid sequence that is the same as part, but not all, of the amino acid sequence of the ADS1, ADS2, or ADS5 polypeptides or one of its functional equivalents. The fragments should comprise at least n consecutive amino acids from the sequence and, depending on the particular sequence, n preferably is 7 or more (for example, 8, 10, 12, 14, 16, 18, 20 or more). Small fragments may form an antigenic determinant.

Preferred polypeptide fragments according to this aspect of the invention are fragments that include a region defined herein as the ADS1, ADS2, or ADS5 adhesion molecule region of the ADS1, ADS2, and ADS5 polypeptides, respectively. These regions are the regions that have been annotated as adhesion molecules.

For the ADS1 polypeptide, this region is considered to extend between residue 250 and residue 365.

For the ADS2 polypeptide, this region is considered to extend between residue 267 and residue 384.

For the ADS5 polypeptide, this region is considered to extend between, at the most, residue 373 and residue 503, and at the least, residue 373 and residue 503.

Variants of this fragment are included as embodiments of this aspect of the invention, provided that these variants possess activity as an adhesion molecule.

In one respect, the term "variant" is meant to include extended or truncated versions of this polypeptide fragment.

For extended variants, it is considered highly likely that the adhesion molecule region of the ADS1, ADS2 and ADS5 polypeptide will fold correctly and show adhesion molecule activity if additional residues C terminal and/or N terminal of these boundaries in the ADS1, ADS2 or ADS5 polypeptide sequences are included in the polypeptide fragment. For example, an additional 5, 10, 20, 30, 40 or even 50 or more amino acid residues from the ADS1, ADS2 or ADS5 polypeptide sequence, or from a homologous sequence, may

be included at either or both the C terminal and/or N terminal of the boundaries of the adhesion molecule regions of the ADS1, ADS2 or ADS5 polypeptide, without prejudicing the ability of the polypeptide fragment to fold correctly and exhibit adhesion molecule activity.

For truncated variants of the ADS1 polypeptide, one or more amino acid residues may be deleted at either or both the C terminus or the N terminus of the adhesion molecule region of the ADS1 polypeptide, although the divalent metal ion binding residues (SER258, SER260 and ASP348), or equivalent residues should be maintained intact; deletions should not extend so far into the polypeptide sequence that any of these residues are deleted.

For truncated variants of the ADS2 polypeptide, one or more amino acid residues may be deleted at either or both the C terminus or the N terminus of the adhesion molecule region of the ADS2 polypeptide, although the divalent metal ion binding residues (SER273, SER275 and ASP365), or equivalent residues should be maintained intact; deletions should not extend so far into the polypeptide sequence that any of these residues are deleted.

For truncated variants of the ADS5 polypeptide, one or more amino acid residues may be deleted at either or both the C terminus or the N terminus of the adhesion molecule region of the ADS5 polypeptide, although the divalent metal ion binding residues (SER378, SER380 and ASP469), or equivalent residues should be maintained intact; deletions should not extend so far into the polypeptide sequence that any of these residues are deleted.

In a second respect, the term "variant" includes homologues of the polypeptide fragments described above, that possess significant sequence homology with the adhesion molecule region of the ADS1 polypeptide and which possess the divalent metal ion binding residues (SER258, SER260 and ASP348), or equivalent residues, provided that said variants retain activity as an adhesion molecule.

The term "variant" also includes homologues of the polypeptide fragments described above, that possess significant sequence homology with the adhesion molecule region of the ADS2 polypeptide and which possess the divalent metal ion binding residues

(SER273, SER275 and ASP365 or equivalent residues), provided that said variants retain activity as an adhesion molecule.

The term "variant" also includes homologues of the polypeptide fragments described above, that possess significant sequence homology with the adhesion molecule region of the ADS5 polypeptide and which possess the divalent metal ion binding residues (SER378, SER380 and ASP469 or equivalent residues), provided that said variants retain activity as an adhesion molecule.

Homologues include those polypeptide molecules that possess greater than 30% identity with the ADS1, ADS2 or ADS5 regions of the ADS1, ADS2 and ADS5 polypeptides, respectively. Percentage identity is as determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for Biotechnology Information; http://www.ncbi.nlm.nih.gov/) [Blosum 62 matrix; gap open penalty=11 and gap extension penalty=1]. Preferably, variant homologues of polypeptide fragments of this aspect of the invention have a degree of sequence identity with the ADS1, ADS2, and ADS5 adhesion molecule regions of the ADS1, ADS2, and ADS5 polypeptides, respectively, of greater than 40%. More preferred variant polypeptides have degrees of identity of greater than 50%, 60%, 70%, 80%, 90%, 95%, 98% or 99%, respectively with the ADS1, ADS2, and ADS5 and adhesion molecule regions of the ADS1, ADS2, or ADS5 polypeptides, provided that said variants retain activity as an adhesion molecule. Variant polypeptides also include homologues of the truncated forms of the polypeptide fragments discussed above, provided that said variants retain activity as an adhesion molecule.

The polypeptide fragments of the first aspect of the invention may be polypeptide fragments that exhibit significant structural homology with the structure of the polypeptide fragment defined by the ADS1, ADS2 or ADS5 adhesion molecule regions, of the ADS1, ADS2 or ADS5 polypeptide sequences, for example, as identified by the Inpharmatica Genome Threader™. Accordingly, polypeptide fragments that are structural homologues of the polypeptide fragments defined by the ADS1, ADS2, or ADS5 adhesion molecule regions of the ADS1, ADS2, and ADS5 polypeptide sequences should

adopt the same fold as that adopted by this polypeptide fragment, as this fold is defined above.

Structural homologues of the polypeptide fragment defined by the ADS1 adhesion molecule region should also retain the divalent metal ion binding residues SER258, SER260 and ASP348, or equivalent residues.

Structural homologues of the polypeptide fragment defined by the ADS2 adhesion molecule region should also retain the divalent metal ion binding residues SER273, SER275 and ASP365, or equivalent residues.

Structural homologues of the polypeptide fragment defined by the ADS5 adhesion molecule region should also retain the divalent metal ion binding residues SER378, SER380 and ASP469, or equivalent residues.

Such fragments may be "free-standing", i.e. not part of or fused to other amino acids or polypeptides, or they may be comprised within a larger polypeptide of which they form a part or region. When comprised within a larger polypeptide, the fragment of the invention most preferably forms a single continuous region. For instance, certain preferred embodiments relate to a fragment having a pre - and/or pro- polypeptide region fused to the amino terminus of the fragment and/or an additional region fused to the carboxyl terminus of the fragment. However, several fragments may be comprised within a single larger polypeptide.

The polypeptides of the present invention or their immunogenic fragments (comprising at least one antigenic determinant) can be used to generate ligands, such as polyclonal or monoclonal antibodies, that are immunospecific for the polypeptides. Such antibodies may be employed to isolate or to identify clones expressing the polypeptides of the invention or to purify the polypeptides by affinity chromatography. The antibodies may also be employed as diagnostic or therapeutic aids, amongst other applications, as will be apparent to the skilled reader.

The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art. As used herein, the term "antibody" refers to intact molecules as well as to

fragments thereof, such as Fab, F(ab')2 and Fv, which are capable of binding to the antigenic determinant in question. Such antibodies thus bind to the polypeptides of the first aspect of the invention.

If polyclonal antibodies are desired, a selected mammal, such as a mouse, rabbit, goat or horse, may be immunised with a polypeptide of the first aspect of the invention. The polypeptide used to immunise the animal can be derived by recombinant DNA technology or can be synthesized chemically. If desired, the polypeptide can be conjugated to a carrier protein. Commonly used carriers to which the polypeptides may be chemically coupled include bovine serum albumin, thyroglobulin and keyhole limpet haemocyanin. The coupled polypeptide is then used to immunise the animal. Serum from the immunised animal is collected and treated according to known procedures, for example by immunoaffinity chromatography.

Monoclonal antibodies to the polypeptides of the first aspect of the invention can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies using hybridoma technology is well known (see, for example, Kohler, G. and Milstein, C., Nature 256: 495-497 (1975); Kozbor et al., Immunology Today 4: 72 (1983); Cole et al., 77-96 in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc. (1985).

Panels of monoclonal antibodies produced against the polypeptides of the first aspect of the invention can be screened for various properties, i.e., for isotype, epitope, affinity, etc. Monoclonal antibodies are particularly useful in purification of the individual polypeptides against which they are directed. Alternatively, genes encoding the monoclonal antibodies of interest may be isolated from hybridomas, for instance by PCR techniques known in the art, and cloned and expressed in appropriate vectors.

Chimeric antibodies, in which non-human variable regions are joined or fused to human constant regions (see, for example, Liu *et al.*, Proc. Natl. Acad. Sci. USA, 84, 3439 (1987)), may also be of use.

The antibody may be modified to make it less immunogenic in an individual, for example by humanisation (see Jones *et al.*, Nature, 321, 522 (1986); Verhoeyen *et al.*, Science, 239: 1534 (1988); Kabat *et al.*, J. Immunol., 147: 1709 (1991); Queen *et al.*, Proc. Natl

Acad. Sci. USA, 86, 10029 (1989); Gorman *et al.*, Proc. Natl Acad. Sci. USA, 88: 34181 (1991); and Hodgson *et al.*, Bio/Technology 9: 421 (1991)). The term "humanised antibody", as used herein, refers to antibody molecules in which the CDR amino acids and selected other amino acids in the variable domains of the heavy and/or light chains of a non-human donor antibody have been substituted in place of the equivalent amino acids in a human antibody. The humanised antibody thus closely resembles a human antibody but has the binding ability of the donor antibody.

In a further alternative, the antibody may be a "bispecific" antibody, that is an antibody having two different antigen binding domains, each domain being directed against a different epitope.

Phage display technology may be utilised to select genes which encode antibodies with binding activities towards the polypeptides of the invention either from repertoires of PCR amplified V-genes of lymphocytes from humans screened for possessing the relevant antibodies, or from naive libraries (McCafferty, J. *et al.*, (1990), Nature 348, 552-554; Marks, J. *et al.*, (1992) Biotechnology 10, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. *et al.*, (1991) Nature 352, 624-628).

Antibodies generated by the above techniques, whether polyclonal or monoclonal, have additional utility in that they may be employed as reagents in immunoassays, radioimmunoassays (RIA) or enzyme-linked immunosorbent assays (ELISA). In these applications, the antibodies can be labelled with an analytically-detectable reagent such as a radioisotope, a fluorescent molecule or an enzyme.

Preferred nucleic acid molecules of the second and third aspects of the invention are those which encode the polypeptide sequences recited in SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6, and functionally equivalent polypeptides, including active fragments of the ADS1, ADS2 and ADS5 polypeptides, such as a fragment including the ADS1, ADS2 or ADS5 adhesion molecule regions of the ADS1, ADS2 and ADS5 polypeptide sequences, or a homologue thereof.

Nucleic acid molecules encompassing these stretches of sequence form a preferred embodiment of this aspect of the invention.

These nucleic acid molecules may be used in the methods and applications described herein. The nucleic acid molecules of the invention preferably comprise at least n consecutive nucleotides from the sequences disclosed herein where, depending on the particular sequence, n is 10 or more (for example, 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

The nucleic acid molecules of the invention also include sequences that are complementary to nucleic acid molecules described above (for example, for antisense or probing purposes).

Nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance cDNA, synthetic DNA or genomic DNA. Such nucleic acid molecules may be obtained by cloning, by chemical synthetic techniques or by a combination thereof. The nucleic acid molecules can be prepared, for example, by chemical synthesis using techniques such as solid phase phosphoramidite chemical synthesis, from genomic or cDNA libraries or by separation from an organism. RNA molecules may generally be generated by the *in vitro* or *in vivo* transcription of DNA sequences.

The nucleic acid molecules may be double-stranded or single-stranded. Single-stranded DNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

The term "nucleic acid molecule" also includes analogues of DNA and RNA, such as those containing modified backbones, and peptide nucleic acids (PNA). The term "PNA", as used herein, refers to an antisense molecule or an anti-gene agent which comprises an oligonucleotide of at least five nucleotides in length linked to a peptide backbone of amino acid residues, which preferably ends in lysine. The terminal lysine confers solubility to the composition. PNAs may be pegylated to extend their lifespan in a cell, where they preferentially bind complementary single stranded DNA and RNA and stop transcript elongation (Nielsen, P.E. *et al.* (1993) Anticancer Drug Des. 8:53-63).

A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:2, or an active fragment thereof, may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:1. These molecules also may have a different sequence which, as a

result of the degeneracy of the genetic code, encodes the polypeptide SEQ ID NO:2, or an active fragment of the ADS1 polypeptide, such as a fragment including the ADS1 adhesion molecule region, or a homologue thereof. The ADS1 adhesion molecule region is considered to extend between residue 250 and residue 365 of the ADS1 polypeptide sequence. In SEQ ID NO:1 the ADS1 adhesion molecule region is thus encoded by a nucleic acid molecule including nucleotide 750 to 1095. Nucleic acid molecules encompassing this stretch of sequence, and homologues of this sequence, form a preferred embodiment of this aspect of the invention.

A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:4, or an active fragment thereof, may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:3. These molecules also may have a different sequence which, as a result of the degeneracy of the genetic code, encodes the polypeptide SEQ ID NO:4, or an active fragment of the ADS2 polypeptide, such as a fragment including the ADS2 adhesion molecule region, or a homologue thereof. The ADS2 adhesion molecule region is considered to extend between residue 267 and residue 384 of the ADS2 polypeptide sequence. In SEQ ID NO:3 the ADS2 adhesion molecule region is encoded by a nucleic acid molecule including nucleotide 801 to nucleotide 1152. Nucleic acid molecules encompassing this stretch of sequence, and homologues of this sequence, form a preferred embodiment of this aspect of the invention.

A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:6, or an active fragment thereof, may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:5. These molecules also may have a different sequence which, as a result of the degeneracy of the genetic code, encodes the polypeptide SEQ ID NO:6, or an active fragment of the ADS5 polypeptide, such as a fragment including the ADS5 adhesion molecule region, or a homologue thereof. The ADS5 adhesion molecule region is considered to extend between residue 373 and residue 503 of the ADS5 polypeptide sequence. In SEQ ID NO:5 the ADS5 adhesion molecule region is encoded by a nucleic acid molecule including nucleotide 1119 to nucleotide 1509. Nucleic acid molecules encompassing this stretch of sequence, and homologues of this sequence, form a preferred embodiment of this aspect of the invention.

Such nucleic acid molecules that encode the polypeptide of SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6 may include, but are not limited to, the coding sequence for the mature polypeptide by itself; the coding sequence for the mature polypeptide and additional coding sequences, such as those encoding a leader or secretory sequence, such as a pro-, pre- or prepro- polypeptide sequence; the coding sequence of the mature polypeptide, with or without the aforementioned additional coding sequences, together with further additional, non-coding sequences, including non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences that play a role in transcription (including termination signals), ribosome binding and mRNA stability. The nucleic acid molecules may also include additional sequences which encode additional amino acids, such as those which provide additional functionalities.

The nucleic acid molecules of the second and third aspects of the invention may also encode the fragments or the functional equivalents of the polypeptides and fragments of the first aspect of the invention.

As discussed above, a preferred fragment of the ADS1 polypeptide is a fragment including the ADS1 adhesion molecule region, or a homologue thereof. The adhesion molecule region is encoded by a nucleic acid molecule including nucleotides 750 to 1095 of SEQ ID NO:1.

A preferred fragment of the ADS2 polypeptide is a fragment including the ADS2 adhesion molecule region, or a homologue thereof. The ADS2 adhesion molecule region is encoded by a nucleic acid molecule including nucleotides 801 to 1152 of SEQ ID NO:3.

A preferred fragment of the ADS5 polypeptide is a fragment including the ADS5 adhesion molecule region, or a homologue thereof. The ADS5 adhesion molecule region is encoded by a nucleic acid molecule including nucleotides 1119 to 1509 of SEQ ID NO:5.

Functionally equivalent nucleic acid molecules according to the invention may be naturally-occurring variants such as a naturally-occurring allelic variant, or the molecules may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the nucleic acid molecule may be made by mutagenesis techniques, including

those applied to nucleic acid molecules, cells or organisms.

Among variants in this regard are variants that differ from the aforementioned nucleic acid molecules by nucleotide substitutions, deletions or insertions. The substitutions, deletions or insertions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or insertions.

The nucleic acid molecules of the invention can also be engineered, using methods generally known in the art, for a variety of reasons, including modifying the cloning, processing, and/or expression of the gene product (the polypeptide). DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides are included as techniques which may be used to engineer the nucleotide sequences. Site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations and so forth.

Nucleic acid molecules which encode a polypeptide of the first aspect of the invention may be ligated to a heterologous sequence so that the combined nucleic acid molecule encodes a fusion protein. Such combined nucleic acid molecules are included within the second or third aspects of the invention. For example, to screen peptide libraries for inhibitors of the activity of the polypeptide, it may be useful to express, using such a combined nucleic acid molecule, a fusion protein that can be recognised by a commercially-available antibody. A fusion protein may also be engineered to contain a cleavage site located between the sequence of the polypeptide of the invention and the sequence of a heterologous protein so that the polypeptide may be cleaved and purified away from the heterologous protein.

The nucleic acid molecules of the invention also include antisense molecules that are partially complementary to nucleic acid molecules encoding polypeptides of the present invention and that therefore hybridize to the encoding nucleic acid molecules (hybridization). Such antisense molecules, such as oligonucleotides, can be designed to recognise, specifically bind to and prevent transcription of a target nucleic acid encoding a polypeptide of the invention, as will be known by those of ordinary skill in the art (see,

for example, Cohen, J.S., Trends in Pharm. Sci., 10, 435 (1989), Okano, J. Neurochem. 56, 560 (1991); O'Connor, J. Neurochem 56, 560 (1991); Lee *et al.*, Nucleic Acids Res 6, 3073 (1979); Cooney *et al.*, Science 241, 456 (1988); Dervan *et al.*, Science 251, 1360 (1991).

The term "hybridization" as used here refers to the association of two nucleic acid molecules with one another by hydrogen bonding. Typically, one molecule will be fixed to a solid support and the other will be free in solution. Then, the two molecules may be placed in contact with one another under conditions that favour hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase molecule to the solid support (Denhardt's reagent or BLOTTO); the concentration of the molecules; use of compounds to increase the rate of association of molecules (dextran sulphate or polyethylene glycol); and the stringency of the washing conditions following hybridization (see Sambrook *et al.* [*supra*]).

The inhibition of hybridization of a completely complementary molecule to a target molecule may be examined using a hybridization assay, as known in the art (see, for example, Sambrook *et al* [*supra*]). A substantially homologous molecule will then compete for and inhibit the binding of a completely homologous molecule to the target molecule under various conditions of stringency, as taught in Wahl, G.M. and S.L. Berger (1987; Methods Enzymol. 152:399-407) and Kimmel, A.R. (1987; Methods Enzymol. 152:507-511).

"Stringency" refers to conditions in a hybridization reaction that favour the association of very similar molecules over association of molecules that differ. High stringency hybridisation conditions are defined as overnight incubation at 42°C in a solution comprising 50% formamide, 5XSSC (150mM NaCl, 15mM trisodium citrate), 50mM sodium phosphate (pH7.6), 5x Denhardts solution, 10% dextran sulphate, and 20 microgram/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1X SSC at approximately 65°C. Low stringency conditions involve the hybridisation reaction being carried out at 35°C (see Sambrook *et al.* [*supra*]). Preferably, the conditions used for hybridization are those of high stringency.

Preferred embodiments of this aspect of the invention are nucleic acid molecules that are at least 70% identical over their entire length to a nucleic acid molecule encoding the ADS1 polypeptide (SEQ ID NO:2), ADS2 polypeptide (SEQ ID NO:4), or ADS5 polypeptide (SEQ ID NO:6), and nucleic acid molecules that are substantially complementary to such nucleic acid molecules. A preferred active fragment is a fragment that includes an ADS1, ADS2, or ADS5 adhesion molecule region of the ADS1, ADS2, and ADS5 polypeptide sequences, resepctively. Accordingly, preferred nucleic acid molecules include those that are at least 70% identical over their entire length to a nucleic acid molecule encoding the adhesion molecule region of the ADS1, ADS2, and ADS5 polypeptide sequence.

Percentage identity, as referred to herein, is as determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for Biotechnology Information; http://www.ncbi.nlm.nih.gov/).

Preferably, a nucleic acid molecule according to this aspect of the invention comprises a region that is at least 80% identical over its entire length to the nucleic acid molecule having the sequence given in SEQ ID NO:1, to a region including nucleotides 750-1095 of this sequence, or a nucleic acid molecule that is complementary to any one of these regions of nucleic acid. In this regard, nucleic acid molecules at least 90%, preferably at least 95%, more preferably at least 98% or 99% identical over their entire length to the same are particularly preferred. Preferred embodiments in this respect are nucleic acid molecules that encode polypeptides which retain substantially the same biological function or activity as the ADS1 polypeptide.

Preferably, a nucleic acid molecule according to this aspect of the invention comprises a region that is at least 80% identical over its entire length to the nucleic acid molecule having the sequence given in SEQ ID NO:3, to a region including nucleotides 801-1152 of this sequence, or a nucleic acid molecule that is complementary to any one of these regions of nucleic acid. In this regard, nucleic acid molecules at least 90%, preferably at least 95%, more preferably at least 98% or 99% identical over their entire length to the same are particularly preferred. Preferred embodiments in this respect are nucleic acid molecules that encode polypeptides which retain substantially the same biological

function or activity as the ADS2 polypeptide.

Preferably, a nucleic acid molecule according to this aspect of the invention comprises a region that is at least 80% identical over its entire length to the nucleic acid molecule having the sequence given in SEQ ID NO:5, to a region including nucleotides 1119 to nucleotide 1509 of this sequence of this sequence, or a nucleic acid molecule that is complementary to any one of these regions of nucleic acid. In this regard, nucleic acid molecules at least 90%, preferably at least 95%, more preferably at least 98% or 99% identical over their entire length to the same are particularly preferred. Preferred embodiments in this respect are nucleic acid molecules that encode polypeptides which retain substantially the same biological function or activity as the ADS5 polypeptide.

The invention also provides a process for detecting a nucleic acid molecule of the invention, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting any such duplexes that are formed.

As discussed additionally below in connection with assays that may be utilised according to the invention, a nucleic acid molecule as described above may be used as a hybridization probe for RNA, cDNA or genomic DNA, in order to isolate full-length cDNAs and genomic clones encoding the ADS1, ADS2 or ADS5 polypeptides and to isolate cDNA and genomic clones of homologous or orthologous genes that have a high sequence similarity to the gene encoding this polypeptide.

In this regard, the following techniques, among others known in the art, may be utilised and are discussed below for purposes of illustration. Methods for DNA sequencing and analysis are well known and are generally available in the art and may, indeed, be used to practice many of the embodiments of the invention discussed herein. Such methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase (US Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proof-reading exonucleases such as those found in the ELONGASE Amplification System marketed by Gibco/BRL (Gaithersburg, MD). Preferably, the sequencing process may be automated using machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), the Peltier

Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

One method for isolating a nucleic acid molecule encoding a polypeptide with an equivalent function to that of the ADS1, ADS2 or ADS5 polypeptides, particularly with an equivalent function to the ADS1, ADS2 or ADS5 adhesion molecule region of the ADS1, ADS2 or ADS5 polypeptides, is to probe a genomic or cDNA library with a natural or artificially-designed probe using standard procedures that are recognised in the art (see, for example, "Current Protocols in Molecular Biology", Ausubel *et al.* (eds). Greene Publishing Association and John Wiley Interscience, New York, 1989,1992). Probes comprising at least 15, preferably at least 30, and more preferably at least 50, contiguous bases that correspond to, or are complementary to, nucleic acid sequences from the appropriate encoding gene (SEQ ID NO:1), particularly a region from nucleotides 750-1095, are particularly useful probes.

Probes comprising at least 15, preferably at least 30, and more preferably at least 50, contiguous bases that correspond to, or are complementary to, nucleic acid sequences from the appropriate encoding gene (SEQ ID NO:3), particularly a region from nucleotides 801-1152 of SEQ ID NO:3, are particularly useful probes.

Probes comprising at least 15, preferably at least 30, and more preferably at least 50, contiguous bases that correspond to, or are complementary to, nucleic acid sequences from the appropriate encoding gene (SEQ ID NO:5), particularly a region from nucleotides 1119 to 1509 of SEQ ID NO:5, are particularly useful probes.

Such probes may be labelled with an analytically-detectable reagent to facilitate their identification. Useful reagents include, but are not limited to, radioisotopes, fluorescent dyes and enzymes that are capable of catalysing the formation of a detectable product. Using these probes, the ordinarily skilled artisan will be capable of isolating complementary copies of genomic DNA, cDNA or RNA polynucleotides encoding proteins of interest from human, mammalian or other animal sources and screening such sources for related sequences, for example, for additional members of the family, type and/or subtype.

In many cases, isolated cDNA sequences will be incomplete, in that the region encoding

the polypeptide will be cut short, normally at the 5' end. Several methods are available to obtain full length cDNAs, or to extend short cDNAs. Such sequences may be extended utilising a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed is based on the method of Rapid Amplification of cDNA Ends (RACE; see, for example, Frohman et al., Proc. Natl. Acad. Sci. USA (1988) 85: 8998-9002). Recent modifications of this technique, exemplified by the MarathonTM technology (Clontech Laboratories Inc.), for example, have significantly simplified the search for longer cDNAs. A slightly different technique, termed "restriction-site" PCR, uses universal primers to retrieve unknown nucleic acid sequence adjacent a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Inverse PCR may also be used to amplify or to extend sequences using divergent primers based on a known region (Triglia, T., *et al.* (1988) Nucleic Acids Res. 16:8186). Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. *et al.* (1991) PCR Methods Applic. 1: 111-119). Another method which may be used to retrieve unknown sequences is that of Parker, J.D. *et al.* (1991); Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PromoterFinderTM libraries to walk genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences that contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

In one embodiment of the invention, the nucleic acid molecules of the present invention may be used for chromosome localisation. In this technique, a nucleic acid molecule is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important step in the confirmatory correlation of those

sequences with the gene-associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationships between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes). This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localised by genetic linkage to a particular genomic region, any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleic acid molecule may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, or affected individuals.

The nucleic acid molecules of the present invention are also valuable for tissue localisation. Such techniques allow the determination of expression patterns of the polypeptide in tissues by detection of the mRNAs that encode them. These techniques include in situ hybridization techniques and nucleotide amplification techniques, such as PCR. Results from these studies provide an indication of the normal functions of the polypeptide in the organism. In addition, comparative studies of the normal expression pattern of mRNAs with that of mRNAs encoded by a mutant gene provide valuable insights into the role of mutant polypeptides in disease. Such inappropriate expression may be of a temporal, spatial or quantitative nature.

The vectors of the present invention comprise nucleic acid molecules of the invention and may be cloning or expression vectors. The host cells of the invention, which may be transformed, transfested or transduced with the vectors of the invention may be prokaryotic or eukaryotic.

The polypeptides of the invention may be prepared in recombinant form by expression of their encoding nucleic acid molecules in vectors contained within a host cell. Such expression methods are well known to those of skill in the art and many are described in detail by Sambrook *et al* (supra) and Fernandez & Hoeffler (1998, eds. "Gene expression

systems. Using nature for the art of expression". Academic Press, San Diego, London, Boston, New York, Sydney, Tokyo, Toronto).

Generally, any system or vector that is suitable to maintain, propagate or express nucleic acid molecules to produce a polypeptide in the required host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those described in Sambrook *et al.*, (supra). Generally, the encoding gene can be placed under the control of a control element such as a promoter, ribosome binding site (for bacterial expression) and, optionally, an operator, so that the DNA sequence encoding the desired polypeptide is transcribed into RNA in the transformed host cell.

Examples of suitable expression systems include, for example, chromosomal, episomal and virus-derived systems, including, for example, vectors derived from: bacterial plasmids, bacteriophage, transposons, yeast episomes, insertion elements, yeast chromosomal elements, viruses such as baculoviruses, papova viruses such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, or combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, including cosmids and phagemids. Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid.

Particularly suitable expression systems include microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (for example, baculovirus); plant cell systems transformed with virus expression vectors (for example, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (for example, Ti or pBR322 plasmids); or animal cell systems. Cell-free translation systems can also be employed to produce the polypeptides of the invention.

Introduction of nucleic acid molecules encoding a polypeptide of the present invention into host cells can be effected by methods described in many standard laboratory manuals, such as Davis et al., Basic Methods in Molecular Biology (1986) and Sambrook

et al., [*supra*]. Particularly suitable methods include calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection (see Sambrook *et al.*, 1989 [*supra*]; Ausubel *et al.*, 1991 [*supra*]; Spector, Goldman & Leinwald, 1998). In eukaryotic cells, expression systems may either be transient (for example, episomal) or permanent (chromosomal integration) according to the needs of the system.

The encoding nucleic acid molecule may or may not include a sequence encoding a control sequence, such as a signal peptide or leader sequence, as desired, for example, for secretion of the translated polypeptide into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals. Leader sequences can be removed by the bacterial host in post-translational processing.

In addition to control sequences, it may be desirable to add regulatory sequences that allow for regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory sequences are those which cause the expression of a gene to be increased or decreased in response to a chemical or physical stimulus, including the presence of a regulatory compound or to various temperature or metabolic conditions. Regulatory sequences are those non-translated regions of the vector, such as enhancers, promoters and 5' and 3' untranslated regions. These interact with host cellular proteins to carry out transcription and translation. Such regulatory sequences may vary in their strength and specificity. Depending on the vector system and host utilised, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript phagemid (Stratagene, LaJolla, CA) or pSportlTM plasmid (Gibco BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (for example, heat shock, RUBISCO and storage protein genes) or from plant viruses (for example, viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line

that contains multiple copies of the sequence, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

An expression vector is constructed so that the particular nucleic acid coding sequence is located in the vector with the appropriate regulatory sequences, the positioning and orientation of the coding sequence with respect to the regulatory sequences being such that the coding sequence is transcribed under the "control" of the regulatory sequences, i.e., RNA polymerase which binds to the DNA molecule at the control sequences transcribes the coding sequence. In some cases it may be necessary to modify the sequence so that it may be attached to the control sequences with the appropriate orientation; i.e., to maintain the reading frame.

The control sequences and other regulatory sequences may be ligated to the nucleic acid coding sequence prior to insertion into a vector. Alternatively, the coding sequence can be cloned directly into an expression vector that already contains the control sequences and an appropriate restriction site.

For long-term, high-yield production of a recombinant polypeptide, stable expression is preferred. For example, cell lines which stably express the polypeptide of interest may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells that successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalised cell lines available from the American Type Culture Collection (ATCC) including, but not limited to, Chinese hamster ovary (CHO), HeLa, baby hamster kidney (BHK), monkey kidney (COS), C127, 3T3, BHK, HEK 293, Bowes melanoma and human hepatocellular carcinoma (for example Hep G2) cells and a number of other cell lines.

In the baculovirus system, the materials for baculovirus/insect cell expression systems are commercially available in kit form from, inter alia, Invitrogen, San Diego CA (the "MaxBac" kit). These techniques are generally known to those skilled in the art and are described fully in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987). Particularly suitable host cells for use in this system include insect cells such as Drosophila S2 and Spodoptera Sf9 cells.

There are many plant cell culture and whole plant genetic expression systems known in the art. Examples of suitable plant cellular genetic expression systems include those described in US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, (1991) Phytochemistry 30, 3861-3863.

In particular, all plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be utilised, so that whole plants are recovered which contain the transferred gene. Practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugar cane, sugar beet, cotton, fruit and other trees, legumes and vegetables.

Examples of particularly preferred bacterial host cells include streptococci, staphylococci, *E. coli*, Streptomyces and *Bacillus subtilis* cells.

Examples of particularly suitable host cells for fungal expression include yeast cells (for example, *S. cerevisiae*) and Aspergillus cells.

Any number of selection systems are known in the art that may be used to recover transformed cell lines. Examples include the herpes simplex virus thymidine kinase (Wigler, M. *et al.* (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. *et al.* (1980) Cell 22:817-23) genes that can be employed in tk- or aprt± cells, respectively.

Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dihydrofolate reductase (DHFR) that confers resistance to methotrexate (Wigler, M. *et al.* (1980) Proc. Natl. Acad. Sci. 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. *et al*

(1981) J. Mol. Biol. 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. Additional selectable genes have been described, examples of which will be clear to those of skill in the art.

Although the presence or absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the relevant sequence is inserted within a marker gene sequence, transformed cells containing the appropriate sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding a polypeptide of the invention under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells that contain a nucleic acid sequence encoding a polypeptide of the invention and which express said polypeptide may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassays, for example, fluorescence activated cell sorting (FACS) or immunoassay techniques (such as the enzyme-linked immunosorbent assay [ELISA] and radioimmunoassay [RIA]), that include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein (see Hampton, R. *et al.* (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN) and Maddox, D.E. *et al.* (1983) J. Exp. Med, 158, 1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labelled hybridization or PCR probes for detecting sequences related to nucleic acid molecules encoding polypeptides of the present invention include oligolabelling, nick translation, end-labelling or PCR amplification using a labelled polynucleotide. Alternatively, the sequences encoding the polypeptide of the invention may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesise RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labelled

nucleotides. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, MI); Promega (Madison WI); and U.S. Biochemical Corp., Cleveland, OH)).

Suitable reporter molecules or labels, which may be used for ease of detection, include radionuclides, enzymes and fluorescent, chemiluminescent or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Nucleic acid molecules according to the present invention may also be used to create transgenic animals, particularly rodent animals. Such transgenic animals form a further aspect of the present invention. This may be done locally by modification of somatic cells, or by germ line therapy to incorporate heritable modifications. Such transgenic animals may be particularly useful in the generation of animal models for drug molecules effective as modulators of the polypeptides of the present invention.

The polypeptide can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulphate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. High performance liquid chromatography is particularly useful for purification. Well known techniques for refolding proteins may be employed to regenerate an active conformation when the polypeptide is denatured during isolation and or purification.

Specialised vector constructions may also be used to facilitate purification of proteins, as desired, by joining sequences encoding the polypeptides of the invention to a nucleotide sequence encoding a polypeptide domain that will facilitate purification of soluble proteins. Examples of such purification-facilitating domains include metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilised metals, protein A domains that allow purification on immobilised immunoglobulin, and the domain utilised in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA) between the purification domain and the polypeptide of the invention may be used to facilitate purification. One such

expression vector provides for expression of a fusion protein containing the polypeptide of the invention fused to several histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification by IMAC (immobilised metal ion affinity chromatography as described in Porath, J. *et al.* (1992) Prot. Exp. Purif. 3: 263-281) while the thioredoxin or enterokinase cleavage site provides a means for purifying the polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D.J. *et al.* (DNA Cell Biol. 199312:441-453).

If the polypeptide is to be expressed for use in screening assays, generally it is preferred that it be produced at the surface of the host cell in which it is expressed. In this event, the host cells may be harvested prior to use in the screening assay, for example using techniques such as fluorescence activated cell sorting (FACS) or immunoaffinity techniques. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the expressed polypeptide. If polypeptide is produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

The polypeptide of the invention can be used to screen libraries of compounds in any of a variety of drug screening techniques. Such compounds may activate (agonise) or inhibit (antagonise) the level of expression of the gene or the activity of the polypeptide of the invention and form a further aspect of the present invention. Preferred compounds are effective to alter the expression of a natural gene which encodes a polypeptide of the first aspect of the invention or to regulate the activity of a polypeptide of the first aspect of the invention.

Agonist or antagonist compounds may be isolated from, for example, cells, cell-free preparations, chemical libraries or natural product mixtures. These agonists or antagonists may be natural or modified substrates, ligands, enzymes, receptors or structural or functional mimetics. For a suitable review of such screening techniques, see Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).

Compounds that are most likely to be good antagonists are molecules that bind to the polypeptide of the invention without inducing the biological effects of the polypeptide upon binding to it. Potential antagonists include small organic molecules, peptides,

polypeptides and antibodies that bind to the polypeptide of the invention and thereby inhibit or extinguish its activity. In this fashion, binding of the polypeptide to normal cellular binding molecules may be inhibited, such that the normal biological activity of the polypeptide is prevented.

The polypeptide of the invention that is employed in such a screening technique may be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. In general, such screening procedures may involve using appropriate cells or cell membranes that express the polypeptide that are contacted with a test compound to observe binding, or stimulation or inhibition of a functional response. The functional response of the cells contacted with the test compound is then compared with control cells that were not contacted with the test compound. Such an assay may assess whether the test compound results in a signal generated by activation of the polypeptide, using an appropriate detection system. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist in the presence of the test compound is observed.

Alternatively, simple binding assays may be used, in which the adherence of a test compound to a surface bearing the polypeptide is detected by means of a label directly or indirectly associated with the test compound or in an assay involving competition with a labelled competitor. In another embodiment, competitive drug screening assays may be used, in which neutralising antibodies that are capable of binding the polypeptide specifically compete with a test compound for binding. In this manner, the antibodies can be used to detect the presence of any test compound that possesses specific binding affinity for the polypeptide.

Assays may also be designed to detect the effect of added test compounds on the production of mRNA encoding the polypeptide in cells. For example, an ELISA may be constructed that measures secreted or cell-associated levels of polypeptide using monoclonal or polyclonal antibodies by standard methods known in the art, and this can be used to search for compounds that may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues. The formation of binding complexes between the polypeptide and the compound being tested may then be

measured.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the polypeptide of interest (see International patent application WO84/03564). In this method, large numbers of different small test compounds are synthesised on a solid substrate, which may then be reacted with the polypeptide of the invention and washed. One way of immobilising the polypeptide is to use non-neutralising antibodies. Bound polypeptide may then be detected using methods that are well known in the art. Purified polypeptide can also be coated directly onto plates for use in the aforementioned drug screening techniques.

The polypeptide of the invention may be used to identify membrane-bound or soluble receptors, through standard receptor binding techniques that are known in the art, such as ligand binding and crosslinking assays in which the polypeptide is labelled with a radioactive isotope, is chemically modified, or is fused to a peptide sequence that facilitates its detection or purification, and incubated with a source of the putative receptor (for example, a composition of cells, cell membranes, cell supernatants, tissue extracts, or bodily fluids). The efficacy of binding may be measured using biophysical techniques such as surface plasmon resonance and spectroscopy. Binding assays may be used for the purification and cloning of the receptor, but may also identify agonists and antagonists of the polypeptide, that compete with the binding of the polypeptide to its receptor. Standard methods for conducting screening assays are well understood in the art.

The invention also includes a screening kit useful in the methods for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, that are described above.

The invention includes the agonists, antagonists, ligands, receptors, substrates and enzymes, and other compounds which modulate the activity or antigenicity of the polypeptide of the invention discovered by the methods that are described above.

The invention also provides pharmaceutical compositions comprising a polypeptide, nucleic acid, ligand or compound of the invention in combination with a suitable pharmaceutical carrier. These compositions may be suitable as therapeutic or diagnostic reagents, as vaccines, or as other immunogenic compositions, as outlined in detail below.

According to the terminology used herein, a composition containing a polypeptide, nucleic acid, ligand or compound [X] is "substantially free of" impurities [herein, Y] when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95%, 98% or even 99% by weight.

The pharmaceutical compositions should preferably comprise a therapeutically effective amount of the polypeptide, nucleic acid molecule, ligand, or compound of the invention. The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent needed to treat, ameliorate, or prevent a targetted disease or condition, or to exhibit a detectable therapeutic or preventative effect. For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, for example, of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

The precise effective amount for a human subject will depend upon the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. This amount can be determined by routine experimentation and is within the judgement of the clinician. Generally, an effective dose will be from 0.01 mg/kg to 50 mg/kg, preferably 0.05 mg/kg to 10 mg/kg. Compositions may be administered individually to a patient or may be administered in combination with other agents, drugs or hormones.

A pharmaceutical composition may also contain a pharmaceutically acceptable carrier, for administration of a therapeutic agent. Such carriers include antibodies and other polypeptides, genes and other therapeutic agents such as liposomes, provided that the carrier does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolised macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid

copolymers and inactive virus particles.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulphates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable carriers is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may additionally contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such compositions. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

The pharmaceutical compositions utilised in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal or transcutaneous applications (for example, see WO98/20734), subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, intravaginal or rectal means. Gene guns or hyposprays may also be used to administer the pharmaceutical compositions of the invention. Typically, the therapeutic compositions may be prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared.

Direct delivery of the compositions will generally be accomplished by injection, subcutaneously, intraperitoneally, intravenously or intramuscularly, or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Dosage treatment may be a single dose schedule or a multiple dose schedule.

If the activity of the polypeptide of the invention is in excess in a particular disease state,

several approaches are available. One approach comprises administering to a subject an inhibitor compound (antagonist) as described above, along with a pharmaceutically acceptable carrier in an amount effective to inhibit the function of the polypeptide, such as by blocking the binding of ligands, substrates, enzymes, receptors, or by inhibiting a second signal, and thereby alleviating the abnormal condition. Preferably, such antagonists are antibodies. Most preferably, such antibodies are chimeric and/or humanised to minimise their immunogenicity, as described previously.

In another approach, soluble forms of the polypeptide that retain binding affinity for the ligand, substrate, enzyme, receptor, in question, may be administered. Typically, the polypeptide may be administered in the form of fragments that retain the relevant portions.

In an alternative approach, expression of the gene encoding the polypeptide can be inhibited using expression blocking techniques, such as the use of antisense nucleic acid molecules (as described above), either internally generated or separately administered. Modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5' or regulatory regions (signal sequence, promoters, enhancers and introns) of the gene encoding the polypeptide. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. *et al.* (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY). The complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes. Such oligonucleotides may be administered or may be generated in situ from expression in vivo.

In addition, expression of the polypeptide of the invention may be prevented by using ribozymes specific to its encoding mRNA sequence. Ribozymes are catalytically active RNAs that can be natural or synthetic (see for example Usman, N, *et al.*, Curr. Opin. Struct. Biol (1996) 6(4), 527-33). Synthetic ribozymes can be designed to specifically

cleave mRNAs at selected positions thereby preventing translation of the mRNAs into functional polypeptide. Ribozymes may be synthesised with a natural ribose phosphate backbone and natural bases, as normally found in RNA molecules. Alternatively the ribozymes may be synthesised with non-natural backbones, for example, 2'-O-methyl RNA, to provide protection from ribonuclease degradation and may contain modified bases.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of non-traditional bases such as inosine, queosine and butosine, as well as acetyl-, methyl-, thio- and similarly modified forms of adenine, cytidine, guanine, thymine and uridine which are not as easily recognised by endogenous endonucleases.

For treating abnormal conditions related to an under-expression of the polypeptide of the invention and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound that activates the polypeptide, i.e., an agonist as described above, to alleviate the abnormal condition. Alternatively, a therapeutic amount of the polypeptide in combination with a suitable pharmaceutical carrier may be administered to restore the relevant physiological balance of polypeptide.

Gene therapy may be employed to effect the endogenous production of the polypeptide by the relevant cells in the subject. Gene therapy is used to treat permanently the inappropriate production of the polypeptide by replacing a defective gene with a corrected therapeutic gene.

Gene therapy of the present invention can occur in vivo or ex vivo. Ex vivo gene therapy requires the isolation and purification of patient cells, the introduction of a therapeutic gene and introduction of the genetically altered cells back into the patient. In contrast, in vivo gene therapy does not require isolation and purification of a patient's cells.

The therapeutic gene is typically "packaged" for administration to a patient. Gene

delivery vehicles may be non-viral, such as liposomes, or replication-deficient viruses, such as adenovirus as described by Berkner, K.L., in Curr. Top. Microbiol. Immunol., 158, 39-66 (1992) or adeno-associated virus (AAV) vectors as described by Muzyczka, N., in Curr. Top. Microbiol. Immunol., 158, 97-129 (1992) and U.S. Patent No. 5,252,479. For example, a nucleic acid molecule encoding a polypeptide of the invention may be engineered for expression in a replication-defective retroviral vector. This expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding the polypeptide, such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells in vivo and expression of the polypeptide in vivo (see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and references cited therein) in Human Molecular Genetics (1996), T Strachan and A P Read, BIOS Scientific Publishers Ltd).

Another approach is the administration of "naked DNA" in which the therapeutic gene is directly injected into the bloodstream or muscle tissue.

In situations in which the polypeptides or nucleic acid molecules of the invention are disease-causing agents, the invention provides that they can be used in vaccines to raise antibodies against the disease causing agent.

Vaccines according to the invention may either be prophylactic (ie. to prevent infection) or therapeutic (ie. to treat disease after infection). Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with pharmaceutically-acceptable carriers as described above, which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, and other pathogens.

Since polypeptides may be broken down in the stomach, vaccines comprising polypeptides are preferably administered parenterally (for instance, subcutaneous,

intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient, and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents.

The vaccine formulations of the invention may be presented in unit-dose or multi-dose containers. For example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

This invention also relates to the use of nucleic acid molecules according to the present invention as diagnostic reagents. Detection of a mutated form of the gene characterised by the nucleic acid molecules of the invention which is associated with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or susceptibility to a disease, which results from under-expression, over-expression or altered spatial or temporal expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.

Nucleic acid molecules for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR, ligase chain reaction (LCR), strand displacement amplification (SDA), or other amplification techniques (see Saiki *et al.*, Nature, 324, 163-166 (1986); Bej, *et al.*, Crit. Rev. Biochem. Molec. Biol., 26, 301-334 (1991); Birkenmeyer et al., J. Virol. Meth., 35, 117-126 (1991); Van Brunt, J., Bio/Technology, 8, 291-294 (1990)) prior to analysis.

In one embodiment, this aspect of the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide according to the invention and comparing said level of expression to a control level, wherein a level that is different to said control level is indicative of disease. The method may comprise the steps of:

a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent

conditions that allow the formation of a hybrid complex between a nucleic acid molecule of the invention and the probe;

b) contacting a control sample with said probe under the same conditions used in step a);

c) and detecting the presence of hybrid complexes in said samples;

wherein detection of levels of the hybrid complex in the patient sample that differ from levels of the hybrid complex in the control sample is indicative of disease.

A further aspect of the invention comprises a diagnostic method comprising the steps of:

a) obtaining a tissue sample from a patient being tested for disease;

b) isolating a nucleic acid molecule according to the invention from said tissue sample; and,

c) diagnosing the patient for disease by detecting the presence of a mutation in the nucleic acid molecule which is associated with disease.

To aid the detection of nucleic acid molecules in the above-described methods, an amplification step, for example using PCR, may be included.

Deletions and insertions can be detected by a change in the size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labelled RNA of the invention or alternatively, labelled antisense DNA sequences of the invention. Perfectly-matched sequences can be distinguished from mismatched duplexes by RNase digestion or by assessing differences in melting temperatures. The presence or absence of the mutation in the patient may be detected by contacting DNA with a nucleic acid probe that hybridises to the DNA under stringent conditions to form a hybrid double-stranded molecule, the hybrid double-stranded molecule having an unhybridised portion of the nucleic acid probe strand at any portion corresponding to a mutation associated with disease; and detecting the presence or absence of an unhybridised portion of the probe strand as an indication of the presence or absence of a disease-associated mutation in the corresponding portion of the DNA strand.

Such diagnostics are particularly useful for prenatal and even neonatal testing.

Point mutations and other sequence differences between the reference gene and "mutant"

genes can be identified by other well-known techniques, such as direct DNA sequencing or single-strand conformational polymorphism, (see Orita *et al.*, Genomics, 5, 874-879 (1989)). For example, a sequencing primer may be used with double-stranded PCR product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabelled nucleotides or by automatic sequencing procedures with fluorescent-tags. Cloned DNA segments may also be used as probes to detect specific DNA segments. The sensitivity of this method is greatly enhanced when combined with PCR. Further, point mutations and other sequence variations, such as polymorphisms, can be detected as described above, for example, through the use of allele-specific oligonucleotides for PCR amplification of sequences that differ by single nucleotides.

DNA sequence differences may also be detected by alterations in the electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (for example, Myers *et al.*, Science (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton et al., Proc. Natl. Acad. Sci. USA (1985) 85: 4397-4401).

In addition to conventional gel electrophoresis and DNA sequencing, mutations such as microdeletions, aneuploidies, translocations, inversions, can also be detected by in situ analysis (see, for example, Keller et al., DNA Probes, 2nd Ed., Stockton Press, New York, N.Y., USA (1993)), that is, DNA or RNA sequences in cells can be analysed for mutations without need for their isolation and/or immobilisation onto a membrane. Fluorescence in situ hybridization (FISH) is presently the most commonly applied method and numerous reviews of FISH have appeared (see, for example, Trachuck *et al.*, Science, 250: 559-562 (1990), and Trask *et al.*, Trends, Genet. 7:149-154 (1991)).

In another embodiment of the invention, an array of oligonucleotide probes comprising a nucleic acid molecule according to the invention can be constructed to conduct efficient screening of genetic variants, mutations and polymorphisms. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic

variability (see for example: M.Chee *et al.*, Science (1996) 274: 610-613).

In one embodiment, the array is prepared and used according to the methods described in PCT application WO95/11995 (Chee *et al*); Lockhart, D. J. *et al.* (1996) Nat. Biotech. 14: 1675-1680); and Schena, M. *et al.* (1996) Proc. Natl. Acad. Sci. 93: 10614-10619). Oligonucleotide pairs may range from two to over one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support. In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application WO95/251116 (Baldeschweiler et al). In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536 or 6144 oligonucleotides, or any other number between two and over one million which lends itself to the efficient use of commercially-available instrumentation.

In addition to the methods discussed above, diseases may be diagnosed by methods comprising determining, from a sample derived from a subject, an abnormally decreased or increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

Assay techniques that can be used to determine levels of a polypeptide of the present invention in a sample derived from a host are well-known to those of skill in the art and are discussed in some detail above (including radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays). This aspect of the invention provides a diagnostic method which comprises the steps of: (a) contacting a ligand as described

above with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

Protocols such as ELISA, RIA, and FACS for measuring polypeptide levels may additionally provide a basis for diagnosing altered or abnormal levels of polypeptide expression. Normal or standard values for polypeptide expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably humans, with antibody to the polypeptide under conditions suitable for complex formation The amount of standard complex formation may be quantified by various methods, such as by photometric means.

Antibodies which specifically bind to a polypeptide of the invention may be used for the diagnosis of conditions or diseases characterised by expression of the polypeptide, or in assays to monitor patients being treated with the polypeptides, nucleic acid molecules, ligands and other compounds of the invention. Antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for the polypeptide include methods that utilise the antibody and a label to detect the polypeptide in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labelled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules known in the art may be used, several of which are described above.

Quantities of polypeptide expressed in subject, control and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease. Diagnostic assays may be used to distinguish between absence, presence, and excess expression of polypeptide and to monitor regulation of polypeptide levels during therapeutic intervention. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials or in monitoring the treatment of an individual patient.

A diagnostic kit of the present invention may comprise:

(a) a nucleic acid molecule of the present invention;

(b) a polypeptide of the present invention; or

(c) a ligand of the present invention.

In one aspect of the invention, a diagnostic kit may comprise a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule according to the invention; a second container containing primers useful for amplifying the nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of disease. The kit may further comprise a third container holding an agent for digesting unhybridised RNA.

In an alternative aspect of the invention, a diagnostic kit may comprise an array of nucleic acid molecules, at least one of which may be a nucleic acid molecule according to the invention.

To detect polypeptide according to the invention, a diagnostic kit may comprise one or more antibodies that bind to a polypeptide according to the invention; and a reagent useful for the detection of a binding reaction between the antibody and the polypeptide.

Such kits will be of use in diagnosing a disease or susceptibility to disease, particularly cardiovascular diseases including atherosclerosis, ischaemia, restenosis, reperfusion injury, sepsis, haematological diseases such as leukaemia, blood clotting disorders, such as thrombosis, cancer including lung, prostate, breast, colorectal and brain tumours, metastasis, inflammatory diseases such as rhinitis, gastrointestinal diseases, including inflammatory bowel disease, ulcerative colitis, Crohn's disease, respiratory diseases including asthma, chronic obstructive pulmonary disease (COPD), respiratory distress syndrome, pulmonary fibrosis, immune disorders, including autoimmune diseases, rheumatoid arthritis, transplant rejection, allergy, liver diseases such as cirrhosis, endocrine diseases, such as diabetes, bone diseases such as osteoporosis, neurological diseases including stroke, multiple sclerosis, spinal cord injury, burns and wound healing, infections, preferably bacterial infection and most preferably *E. coli* infection.

Various aspects and embodiments of the present invention will now be described in more detail by way of example, with particular reference to the ADS1, ADS2 and ADS5 polypeptides.

It will be appreciated that modification of detail may be made without departing from the

scope of the invention.

Brief description of the Figures

Figure 1: This is the front end of the Biopendium™ Target Mining Interface. A search of the database is initiated using the PDB code "1LFA:A".

Figure 2A: A selection is shown of the Inpharmatica Genome Threader results for the search using 1LFA:A. The arrow indicates leukocyte integrin, a typical adhesion molecule.

Figure 2B: A selection is shown of the Inpharmatica Genome Threader results for the search using 1LFA:A. The arrow indicates AAC74854.1 (ADS1).

Figure 2C: Full list of forward PSI-BLAST results for the search using 1LFA:A. AAC74854.1 (ADS1) is not identified.

Figure 3: The Redundant Sequence Display results page for AAC74854.1 (ADS1).

Figure 4: PFAM search results for AAC74854.1 (ADS1).

Figure 5: NCBI protein report for AAC74854.1 (ADS1).

Figure 6A: This is the front end of the Biopendium™ database. A search of the database is initiated using AAC74854.1 (ADS1), as the query sequence.

Figure 6B: A selection of the Inpharmatica Genome Threader results of search using AAC74854.1 (ADS1), as the query sequence. The arrow points to 1LFA:A.

Figure 6C: A selection of the reverse-maximised PSI-BLAST results obtained using AAC74854.1 (ADS1), as the query sequence.

Figure 7: AlEye sequence alignment of BAA15585.1 (AAC74854.1 (ADS1)) and 1LFA:A.

Figure 8A: LigEye for 1LFA: A that illustrates the sites of interaction of the bound metal ion required for adhesion activity with the metal binding ligands of the MIDAS motif of *Homo Sapiens* Leukocyte Function Antigen 1, 1LFA: A

Figure 8B: iRasMol view of 1LFA: A, *Homo Sapiens* Leukocyte Function Antigen 1. The coloured balls represent the amino acids in *Homo Sapiens* Leukocyte Function

Antigen 1 that are involved in the MIDAS motif and that are conserved in AAC74854.1 (ADS1).

Figure 9: This is the front end of the Biopendium™ Target Mining Interface. A search of the database is initiated using the PDB code "1AOX:A".

Figure 10A: A selection is shown of the Inpharmatica Genome Threader results for the search using 1AOX:A. The arrow indicates leukocyte integrin, a typical adhesion molecule.

Figure 10B: A selection is shown of the Inpharmatica Genome Threader results for the search using 1AOX:A. The arrow indicates AAC76768.1 (ADS2).

Figure 10C: Full list of forward PSI-BLAST results for the search using 1AOX:A. AAC76768.1 (ADS2) is not identified.

Figure 11: The Redundant Sequence Display results page for AAC76768.1 (ADS2).

Figure 12: PFAM search results for AAC76768.1 (ADS2).

Figure 13: NCBI protein report for AAC76768.1 (ADS2).

Figure 14A: This is the front end of the Biopendium™ database. A search of the database is initiated using AAC76768.1 (ADS2), as the query sequence.

Figure 14B: A selection of the Inpharmatica Genome Threader results of search using AAC76768.1 (ADS2), as the query sequence. The arrow points to 1AOX:A.

Figure 14C: A selection of the reverse-maximised PSI-BLAST results obtained using AAC76768.1 (ADS2), as the query sequence.

Figure 15: AlEye sequence alignment of P03818 (AAC76768.1 (ADS2)) and 1AOX:A.

Figure 16A: LigEye for 1AOX: A that illustrates the sites of interaction of the bound metal ion required for adhesion activity with the metal binding ligands of the MIDAS motif of *Homo Sapiens* Integrin Alpha 2 / Beta 1, 1AOX:A

Figure 16B: iRasMol view of 1AOX:A, *Homo Sapiens* Integrin Alpha 2 / Beta 1. The coloured balls represent the amino acids in *Homo Sapiens* Integrin Alpha 2 / Beta 1 that are involved in the MIDAS motif and that are conserved in AAC76768.1 (ADS2).

Figure 17: This is the front end of the Biopendium™ Target Mining Interface. A search of the database is initiated using the PDB code "1JLM".

Figure 18A: A selection is shown of the Inpharmatica Genome Threader results for the search using 1JLM. The arrow indicates leukocyte integrin, a typical adhesion molecule.

Figure 18B: A selection is shown of the Inpharmatica Genome Threader results for the search using 1JLM. The arrow indicates P10155 (ADS5).

Figure 18C: Full list of forward PSI-BLAST results for the search using 1JLM. P10155 (ADS5) is not identified.

Figure 19: The Redundant Sequence Display results page for P10155 (ADS5).

Figure 20: PFAM search results for P10155 (ADS5).

Figure 21: NCBI protein report for P10155 (ADS5).

Figure 22A: This is the front end of the Biopendium™ database. A search of the database is initiated using P10155 (ADS5), as the query sequence.

Figure 22B: A selection of the Inpharmatica Genome Threader results of search using P10155 (ADS5), as the query sequence. The arrow points to 1JLM.

Figure 22C: A selection of the reverse-maximised PSI-BLAST results obtained using P10155 (ADS5), as the query sequence.

Figure 23: AlEye sequence alignment of P10155 (ADS5) and 1JLM.

Figure 24A: LigEye for 1JLM that illustrates the sites of interaction of the bound metal ion required for adhesion activity with the metal binding ligands of the MIDAS motif of *Homo Sapiens* Integrin CR3, 1JLM

Figure 24B: iRasMol view of 1JLM, *Homo Sapiens* Integrin CR3. The coloured balls represent the amino acids in *Homo Sapiens* Integrin CR3 that are involved in the MIDAS motif and that are partly conserved in P10155 (ADS5).

Figure 25: AlEye sequence alignment of P10155, *Homo Sapiens* Ro60 (ADS5), and the *Mus musculus* (AAF19049.1), *Xenopus Laevis* (AAC38001.1) and *Caenorhabditis elegans* (CAA98241.1) Ro60 homologs.

Examples

Example 1:AAC74854.1 (ADS1)

In order to initiate a search for novel, distantly related adhesion molecules, an archetypal family member is chosen, the I-domain from *Homo Sapiens* Leukocyte Function Antigen 1. More specifically, the search is initiated using a structure from the Protein Data Bank (PDB) which is operated by the Research Collaboratory for Structural Bioinformatics.

The structure chosen is the I-domain from *Homo Sapiens* Leukocyte Function Antigen 1 (PDB code 1LFA:A; see Figure 1).

A search of the Biopendium™ (using the Target Mining Interface) for relatives of 1LFA:A takes place and returns 2729 Genome Threader results. The 2729 Genome Threader results include examples of typical adhesion molecules, such as leukocyte integrin alpha chain (see arrow in Figure 2A).

Among the known adhesion molecules appears a protein of apparently unknown function, AAC74854.1 (ADS1; see arrow in figure 2B). The Inpharmatica Genome Threader has identified a sequence, AAC74854.1 (ADS1), as having a structure similar to *Homo Sapiens* Leukocyte Function Antigen 1, an adhesion molecule. The possession of a structure similar to an adhesion molecule suggests that AAC74854.1 (ADS1) functions as an adhesion molecule. The Genome Threader identifies this with 95% confidence.

The search of the Biopendium™ (using the Target Mining Interface) for relatives of 1LFA:A also returns 630 Forward PSI-Blast results. Forward PSI-Blast (see figure 2C) is unable to identify this relationship; only the Inpharmatica Genome Threader is able to identify AAC74854.1 (ADS1) as an adhesion molecule.

In order to assess what is known in the public domain databases about AAC74854.1 (ADS1) the Redundant Sequence Display Page (Figure 3) is viewed. There are two other sequences in the public domain which have an identical sequence to AAC74854.1 and hence appear on this page, P76235 and BAA15585.1. There are no associated PROSITE or PRINTS hits for AAC74854.1 (ADS1). PROSITE and PRINTS are databases that help to describe proteins of similar families. Returning no hits from both databases means that

AAC74854.1 (ADS1) is unidentifiable as an adhesion molecule using PROSITE or PRINTS. The redundant sequence display also shows any predicted features of AAC50543.1 (CCS5). These include potential coiled coil and low complexity regions in the sequence.

In order to identify if any other public domain annotation vehicle is able to annotate AAC74854.1 (ADS1) as an adhesion molecule, the AAC74854.1 (ADS1) protein sequence is searched against the PFAM database (Protein Family Database of Alignment and hidden Markov models) (see Figure 4). The results identifies one PFAM-B match to AAC74854.1, however PFAM-B matches confer no functional annotation, only sequence similarity to other functionally unannotated proteins. Thus PFAM does not identify AAC74854.1 (ADS1) as an adhesion molecule.

The National Center for Biotechnology Information (NCBI) Genbank protein database is then viewed to examine if there is any further information that is known in the public domain relating to AAC74854.1 (ADS1). This is the U.S. public domain database for protein and gene sequence deposition (Figure 5). AAC74854.1 (ADS1) is an *Escherichia Coli* sequence, its Genbank protein ID is AAC74854.1 and it is 427 amino acids in length. AAC74854.1 (ADS1) was cloned by a group of scientists at the University of Wisconsin, U.S.A. The entry identifies AAC74854.1 (ADS1) as a hypothetical protein. The public domain information for this gene does not annotate it as an adhesion molecule.

Therefore, it can be concluded that using all public domain annotation tools, AAC74854.1 (ADS1) may not be annotated as an adhesion molecule. Only the Inpharmatica Genome Threader is able to annotate this protein as an adhesion molecule.

The reverse search is now carried out. AAC74854.1 (ADS1) is now used as the query sequence in the Biopendium™ (see figure 6A). The Inpharmatica Genome Threader identifies AAC74854.1 (ADS1) as having a structure that is the same as *Homo Sapiens* Leukocyte Function Antigen 1 with 95% confidence (see arrow in figure 6B). *Homo Sapiens* Leukocyte Function Antigen 1 (1LFA) was the original query sequence. Positive iterations of PSI-Blast do not return this result (Figure 6C). It is only the Inpharmatica Genome Threader that is able to identify this relationship.

The *Homo Sapiens* Leukocyte Function Antigen 1 sequence is chosen against which to view the sequence alignment of BAA15585.1 (AAC74854.1 (ADS1)). Viewing the AlEye alignment (Figure 7) of the query protein against the protein identified as being of a similar structure helps to visualize the areas of homology.

The Leukocyte Function Antigen 1 I domain requires a bound metal ion in order to function. The metal ion forms a Metal Ion-Dependent Adhesion Site (MIDAS) which is characterised by a MIDAS motif consisting of the conserved metal liganding residues. The MIDAS motif in 1LFA:A consists of ASP10, SER12, SER14, THR79 and ASP112 all these residues are conserved in BAA15585.1 (AAC74854.1 (ADS1)) as ASP256, SER258, SER260, THR315 and ASP346 respectively. The two serines and ASP112 are the metal ion ligands. This indicates that AAC74854.1 (ADS1) is an adhesion molecule similar to Leukocyte Function Antigen 1.

In order to ensure that the protein identified is in fact a relative of the query sequence, the visualization programs "LigEye" (Figure 8A) and "iRasmol" (Figure 8B) are used. These visualization tools identify the metal binding site of known protein structures by indicating the amino acids with which known metal ions or small molecule inhibitors interact at the active site. These interactions are through either a direct hydrogen bond or through hydrophobic interactions. In this manner one can see if the active site fold/structure is conserved between the identified homologue and the chosen protein of known structure.

Since the structure of *Homo Sapiens* Leukocyte Function Antigen 1 is known (1LFA), this is chosen to illustrate the MIDAS motif (Figure 8B). ASP10, SER12, SER14, THR79 and ASP112 of 1LFA:A align with ASP256, SER258, SER260, THR315 and ASP346 of AAC74854.1. This indicates that indeed as predicted by the Inpharmatica Genome Threader, AAC74854.1 (ADS1) folds in a similar manner to *Homo Sapiens* Leukocyte Function Antigen 1 and as such is identified as an adhesion molecule.

Example 2:AAC76768.1 (ADS2)

In order to initiate a search for novel, distantly related adhesion molecules, an archetypal family member is chosen, the I-domain from *Homo Sapiens* Integrin Alpha 2 / Beta 1.

More specifically, the search is initiated using a structure from the Protein Data Bank (PDB) which is operated by the Research Collaboratory for Structural Bioinformatics.

The structure chosen is the I-domain from *Homo Sapiens* Integrin Alpha 2 / Beta 1 (PDB code 1AOX; see Figure 9).

A search of the Biopendium™ (using the Target Mining Interface) for relatives of 1AOX takes place and returns 2394 Genome Threader results. The 2394 Genome Threader results include examples of typical adhesion molecules, such as Integrin alpha 11 (see arrow in Figure 10A).

Among the known adhesion molecules appears a protein of apparently unknown function, AAC76768.1 (ADS2; see arrow in figure 10B). The Inpharmatica Genome Threader has identified a sequence, AAC76768.1 (ADS2), as having a structure similar to *Homo Sapiens* Integrin Alpha 2 / Beta 1, an adhesion molecule. The possession of a structure similar to an adhesion molecule suggests that AAC76768.1 (ADS2) functions as an adhesion molecule. The Genome Threader identifies this with 100% confidence.

The search of the BiopendiumTM (using the Target Mining Interface) for homologues of 1FBL also returns 24 Reverse PSI-Blast results. The Inpharmatica Reverse PSI-Blast identifies AAC76768.1 (ADS2) as being related in sequence to *Homo Sapiens* Integrin Alpha 2 / Beta 1, detected in the -4 iteration (see Figure 10B, circled). The possession of a sequence related to an adhesion molecule suggests that AAC76768.1 (ADS2) functions as an adhesion molecule. This second proprietary method result consolidates the *Homo Sapiens* Integrin Alpha 2 / Beta 1 structural relationship demonstrated with Genome Threader.

The search of the Biopendium™ (using the Target Mining Interface) for relatives of 1AOX also returns 608 Forward PSI-Blast results. Forward PSI-Blast (see figure 10C) is unable to identify this relationship; only the Inpharmatica Genome Threader is able to identify AAC76768.1 (ADS2) as an adhesion molecule.

In order to assess what is known in the public domain databases about AAC76768.1 (ADS2) the Redundant Sequence Display Page (Figure 11) is viewed. There are two other sequences in the public domain which have the identical sequence to AAC76768.1

and hence appear on this page, AAA62097.1 and P03818. P03818 is identical in sequence and length to AAC76768.1. There are no associated PROSITE or PRINTS hits for AAC76768.1 (ADS2). PROSITE and PRINTS are databases that help to describe proteins of similar families. Returning no hits from both databases means that AAC76768.1 (ADS2) is unidentifiable as an adhesion molecule using PROSITE or PRINTS. The redundant sequence display also shows any predicted features of AAC50543.1 (CCS5). These include a potential coiled coil in the sequence.

In order to identify if any other public domain annotation vehicle is able to annotate AAC76768.1 (ADS2) as an adhesion molecule, the AAC76768.1 (ADS2) protein sequence is searched against the PFAM database (Protein Family Database of Alignment and hidden Markov models) (see Figure 12). The results identifies one PFAM-B match to AAC76768.1, however PFAM-B matches confer no functional annotation, only sequence similarity to other functionally unannotated proteins. Thus PFAM does not identify AAC76768.1 (ADS2) as an adhesion molecule.

The National Center for Biotechnology Information (NCBI) Genbank protein database is then viewed to examine if there is any further information that is known in the public domain relating to AAC76768.1 (ADS2). This is the U.S. public domain database for protein and gene sequence deposition (Figure 13). AAC76768.1 (ADS2) is an *Escherichia Coli* sequence, its Genbank protein ID is AAC76768.1 and it is 427 amino acids in length. AAC76768.1 (ADS2) was cloned by a group of scientists at the University of Wisconssin, USA. The entry identifies AAC76768.1 (ADS2) as a hypothetical protein. The public domain information for this gene does not annotate it as an adhesion molecule.

Therefore, it can be concluded that using all public domain annotation tools, AAC76768.1 (ADS2) may not be annotated as an adhesion molecule. Only the Inpharmatica Genome Threader is able to annotate this protein as an adhesion molecule.

The reverse search is now carried out. AAC76768.1 (ADS2) is now used as the query sequence in the Biopendium™ (see figure 14A). The Inpharmatica Genome Threader identifies AAC76768.1 (ADS2) as having a structure that is the same as *Homo Sapiens* Integrin Alpha 2 / Beta 1 with 100% confidence (see arrow in figure 14B). *Homo Sapiens*

Integrin Alpha 2 / Beta 1 (1AOX) was the original query sequence. The first 3 iterations of positive PSI-Blast do not return this result (Figure 14C), adhesion molecules are only detected at and above iteration 4. It is only the Inpharmatica Genome Threader that is able to identify this relationship.

The *Homo Sapiens* Integrin Alpha 2 / Beta 1 sequence is chosen against which to view the sequence alignment of AAC76768.1 (ADS2). Viewing the AlEye alignment (Figure 15) of the query protein against the protein identified as being of a similar structure helps to visualize the areas of homology.

The Integrin Alpha 2 / Beta 1 I domain requires a bound metal ion in order to function. The metal ion forms a Metal Ion-Dependent Adhesion Site (MIDAS) which is characterised by a MIDAS motif consisting of the conserved metal liganding residues. The MIDAS motif in 1AOX consists of ASP13, SER15, SER17, THR83 and ASP116 all these residues are conserved in AAC76768.1 (ADS2) as ASP271, SER273, SER275, THR337 and ASP365 respectively. The two serines and ASP116 are the metal ion ligands. This indicates that AAC76768.1 (ADS2) is an adhesion molecule similar to Integrin Alpha 2 / Beta 1.

In order to ensure that the protein identified is in fact a relative of the query sequence, the visualization programs "LigEye" (Figure 16A) and "iRasmol" (Figure 16B) are used. These visualization tools identify the metal binding site of known protein structures by indicating the amino acids with which known metal ions or small molecule inhibitors interact at the active site. These interactions are through either a direct hydrogen bond or through hydrophobic interactions. In this manner one can see if the active site fold/structure is conserved between the identified homologue and the chosen protein of known structure.

Since the structure of *Homo Sapiens* Integrin Alpha 2 / Beta 1 is known (1AOX), this is chosen to illustrate the MIDAS motif (Figure 16B). ASP13, SER15, SER17, THR83 and ASP116 of 1AOX align with ASP271, SER273, SER275, THR337 and ASP365 of AAC76768.1. This indicates that indeed as predicted by the Inpharmatica Genome Threader, AAC76768.1 (ADS2) folds in a similar manner to *Homo Sapiens* Integrin Alpha 2 / Beta 1 and as such is identified as an adhesion molecule.

Example 3:Ro60 (ADS5)

In order to initiate a search for novel, distantly related adhesion molecules, an archetypal family member is chosen, the I-domain from *Homo Sapiens* Integrin CR3. More specifically, the search is initiated using a structure from the Protein Data Bank (PDB) which is operated by the Research Collaboratory for Structural Bioinformatics.

The structure chosen is the I-domain from *Homo Sapiens* Integrin CR3 (PDB code 1JLM; see Figure 17).

A search of the Biopendium™ (using the Target Mining Interface) for relatives of 1JLM takes place and returns 2925 Genome Threader results. The 2925 Genome Threader results include examples of typical adhesion molecules, such as leukocyte integrin (see arrow in Figure 18A).

Among the known adhesion molecules appears a protein of apparently unknown function, Ro60 (ADS5; see arrow in Figure 18B). The Inpharmatica Genome Threader has identified a sequence, Ro60 (ADS5), as having a structure similar to *Homo Sapiens* Integrin CR3, an adhesion molecule. The possession of a structure similar to an adhesion molecule suggests that Ro60 (ADS5) functions as an adhesion molecule. The Genome Threader identifies this with 70% confidence.

The search of the Biopendium™ (using the Target Mining Interface) for relatives of 1JLM also returns 626 Forward PSI-Blast results. Forward PSI-Blast (see Figure 18C) is unable to identify this relationship; only the Inpharmatica Genome Threader is able to identify Ro60 (ADS5) as an adhesion molecule.

In order to assess what is known in the public domain databases about Ro60 (ADS5) the Redundant Sequence Display Page (Figure 19) is viewed. There are no associated PROSITE or PRINTS hits for Ro60 (ADS5). PROSITE and PRINTS are databases that help to describe proteins of similar families. Returning no hits from both databases means that Ro60 (ADS5) is unidentifiable as an adhesion molecule using PROSITE or PRINTS. The redundant sequence display also shows any predicted features of Ro60 (ADS5). These include a potential coiled coil region at the start of the sequence and a transmembrane region. Although transmembrane regions are not predictive of adhesion

molecules, they are a common characteristic of adhesion molecules. Thus the possession of a transmembrane region consolidates the Inpharmatica Genome Threader annotation of Ro60 (ADS5) as an adhesion molecule.

In order to identify if any other public domain annotation vehicle is able to annotate Ro60 (ADS5) as an adhesion molecule, the Ro60 (ADS5) protein sequence is searched against the PFAM database (Protein Family Database of Alignment and hidden Markov models) (see Figure 20). The results identifies two PFAM-B matches to Ro60, however PFAM-B matches confer no functional annotation, only sequence similarity to other functionally unannotated proteins. Thus PFAM does not identify Ro60 (ADS5) as an adhesion molecule.

The National Center for Biotechnology Information (NCBI) Genbank protein database is then viewed to examine if there is any further information that is known in the public domain relating to Ro60 (ADS5). This is the U.S. public domain database for protein and gene sequence deposition (Figure 21). Ro60 (ADS5) is a *Homo Sapiens* sequence, its SWISS-PROT protein ID is P10155 and it is 538 amino acids in length. Ro60 (ADS5) was cloned by a group of scientists at the W.M. Keck Autoimmune Disease Center, California. The entry identifies Ro60 (ADS5) as a RNA binding protein associated with the autoimmune disease: Sjogren's Syndrome. The public domain information for this gene does not annotate it as an adhesion molecule.

Therefore, it can be concluded that using all public domain annotation tools, Ro60 (ADS5) may not be annotated as an adhesion molecule. Only the Inpharmatica Genome Threader is able to annotate this protein as an adhesion molecule.

The reverse search is now carried out. P10155 (ADS5) is now used as the query sequence in the Biopendium™ (see Figure 22A). The Inpharmatica Genome Threader identifies Ro60 (ADS5) as having a structure that is the same as *Homo Sapiens* Integrin CR3 with 70% confidence (see arrow in Figure 22B). *Homo Sapiens* Integrin CR3 (1JLM) was the original query sequence. Positive iterations of PSI-Blast do not return this result (Figure 22C). It is only the Inpharmatica Genome Threader that is able to identify this relationship.

The *Homo Sapiens* Integrin CR3 sequence is chosen against which to view the sequence alignment of Ro60 (ADS5). Viewing the AlEye alignment (Figure 23) of the query protein against the protein identified as being of a similar structure helps to visualize the areas of homology.

The Integrin CR3 I domain requires a bound metal ion in order to function. The metal ion forms a Metal Ion-Dependent Adhesion Site (MIDAS) which is characterised by a MIDAS motif consisting of the conserved metal liganding residues. The MIDAS motif in 1JLM consists of ASP9, SER11, SER13, THR78 and ASP111 all these residues except for THR78 are conserved in Ro60 (ADS5) as ASP376, SER378, SER380, ILE440 and ASP469 respectively. The two serines and ASP111 are the metal ion ligands. This indicates that Ro60 (ADS5) is an adhesion molecule similar to Integrin CR3.

In order to ensure that the protein identified is in fact a relative of the query sequence, the visualization programs "LigEye" (Figure 24A) and "iRasmol" (Figure 24B) are used. These visualization tools identify the metal binding site of known protein structures by indicating the amino acids with which known metal ions or small molecule inhibitors interact at the active site. These interactions are through either a direct hydrogen bond or through hydrophobic interactions. In this manner one can see if the active site fold/structure is conserved between the identified homologue and the chosen protein of known structure.

Since the structure of *Homo Sapiens* Integrin CR3 is known (1JLM), this is chosen to illustrate the MIDAS motif (Figure 63B). ASP9, SER11, SER13, THR78 and ASP111 of 1JLM align with ASP376, SER378, SER380, ILE440 and ASP469 of Ro60. This indicates that indeed as predicted by the Inpharmatica Genome Threader, Ro60 (ADS5) folds in a similar manner to *Homo Sapiens* Integrin CR3 and as such is identified as an adhesion molecule.

Reverse-maximised PSI-BLAST of Ro60 (ADS5) identifies *Mus musculus, Xenopus Laevis* and *Caenorhabditis elegans* homologs of Ro60 (ADS5) called AAF19049.1, AAC38001.1, and CAA98241.1 respectively. AAF19049.1 has 90.0% sequence identity to P10155 (*Homo sapiens* Ro60; ADS5), see Figure 22C. AAC38001.1 has 76.0%

sequence identity to P10155 (*Homo sapiens* Ro60; ADS5), see Figure 22C. CAA98241.1 has 36.0% sequence identity to P10155 (*Homo sapiens* Ro60; ADS5), see Figure 22C.

P10155 (*Homo sapiens* Ro60; ADS5), called AAF19049.1, AAC38001.1, and CAA98241.1 are aligned and viewed in AlEye (Figure 25). AlEye reveals that the predicted metal binding residues SER378, SER380 and ASP469 of P10155 (*Homo sapiens* Ro60;ADS5), are conserved in the *Mus musculus* (AAF19049.1), *Xenopus Laevis* (AAC38001.1) and *Caenorhabditis elegans* (CAA98241.1) DICE-1 homologs. Another predicted MIDAS residue of (*Homo sapiens* Ro60;ADS5), ASP376, is also conserved. Residues, which are essential for the function of a protein, will be conserved in homologs of that protein. Thus the conservation of SER378, SER380 and ASP469 (residues which would be essential for the function of the I-domain) and ASP376 in the *Mus musculus* (AAF19049.1), *Xenopus Laevis* (AAC38001.1) and *Caenorhabditis elegans* (CAA98241.1) DICE-1 homologs strongly supports the annotation of P10155 (*Homo sapiens* Ro60; ADS5) as an adhesion molecule.

Sequence Listing

SEQ ID NO:1 (Nucleotide coding sequence for AAC74854.1 (ADS1) protein)

```
   1 atgacctggt ttattgaccg gcgtctgaac ggcaaaaaca aaagcatggt gaatcgccag
  61 cgtttttttac gccgttataa agcgcaaatt aaacagtcga tctccgaggc cattaataag
 121 cgttcggtga ctgacgtcga cagcggcgaa tccgtatcca ttcccacgga agatattagc
 181 gaaccgatgt ttcatcaggg gcgtggcggt ctgcgccacc gcgtgcatcc gggcaatgac
 241 catttcgtcc agaacgaccg aattgaacgt ccccagggtg gcggcggagg ttccggcagt
 301 ggtcagggcc aggccagcca ggatggtgaa ggtcaggatg aatttgtctt tcagatttcg
 361 aaagatgagt atcttgatct gctctttgaa gatttggcct taccgaatct gaaacaaaac
 421 caacaacgcc agctgaccga atataaaacg catcgggcgg gttataccgc taacggcgtt
 481 ccggccaata tcagcgttgt gcgttcattg cagaactcac tggcgcgacg cacagccatg
 541 acggcaggca agcggcggga acttcatgca ctggaagaga atttggccat catcagcaac
 601 agtgaacctg cgcaactgct ggaagaggaa cgtctgcgca aagaaattgc agaattacgt
 661 gccaaaattg aacgcgtccc ttttattgac accttcgatt tacgttacaa gaactacgag
 721 aagcggcccg atccctccag ccaggcagtg atgttttgcc tgatggacgt ttccggttca
 781 atggatcaat ccactaaaga tatggctaag cgtttttata ttctgctgta tctgttcctc
 841 agcagaacgt ataagaacgt ggaagtcgta tacatccgcc atcataccca ggcgaaagaa
 901 gtcgatgaac atgagttttt ctactcgcag gaaacaggcg gcaccattgt ttccagcgcc
 961 ctgaaactga tggatgaggt agtgaaagag cgttataacc cggcacagtg gaatatttac
1021 gctgcacaag catcggacgg cgataactgg gccgatgact ctccgctttg ccatgaaatc
1081 ctggcgaaaa aattattacc tgttgttcgt tattacagct atatcgaaat tacccgtcgt
1141 gcacatcaga cattgtggcg agaatatgag catctgcaat ctactttcga caactttgcg
1201 atgcagcaca tccgcgacca ggatgatatt tatccggtgt tccgtgaact gtttcataaa
1261 caaaatgcaa cagctaaagg ctaa
```

SEQ ID NO:2 (Protein AAC74854.1; ADS1)

```
   1 mtwfidrrln gknksmvnrq rflrrykaqi kqsiseaink rsvtdvdsge svsiptedis
  61 epmfhqgrgg lrhrvhpgnd hfvqndrier pqgggggsgs gqgqasqdge gqdefvfqis
 121 kdeyldllfe dlalpnlkqn qqrqlteykt hragytangv panisvvrsl qnslarrtam
 181 tagkrrelha leenlaiisn sepaqlleee rlrkeiaelr akiervpfid tfdlryknye
 241 krpdpssqav mfclmdvsgs mdqstkdmak rfyillylfl srtyknvevv yirhhtqake
```

301 vdeheffysq etggtivssa lklmdevvke rynpaqwniy aaqasdgdnw addsplchei

361 lakkllpvvr yysyieitrr ahqtlwreye hlqstfdnfa mqhirdqddi ypvfrelfhk

421 qnatakg

SEQ ID NO:3 (the nucleotide coding sequence for AAC76768.1 (ADS2) protein)

1 gtgcgcagtc ggctgaaaga tgcccgagtc ccgccggaac tcaccgaaga ggtgatgtgc

61 tatcagcaaa gccagctcct ctccacgcca cagtttattg tgcagctacc acagatcctg

121 gacttactgc atcgtctgaa ttctccatgg gcagaacaag cccgacagtt ggttgatgct

181 aacagcacga tcacttcagc gttacacacg ctttttctcc agcgttggcg tttaagtctg

241 atcgtgcaag caacgacgtt aaatcaacag ctattagaag aagaacgcga acaactgttg

301 agtgaagttc aggaacgcat gacgctgagc ggacaacttg aaccgattct cgcagataac

361 aatactgcag ctggtcgtct gtgggatatg agcgccggtc agcttaaacg tggcgactat

421 cagttgattg tgaaatacgg tgaatttctt aacgaacagc cggaactgaa acgcctggca

481 gagcagctgg ggcgttctcg ggaagccaaa tcaataccgc gcaacgatgc gcagatggaa

541 accttccgca ccatggtgcg cgaaccggcg acggttcctg agcaggttga tggtctgcaa

601 caaagcgatg atattttacg tctcctgccg ccagaactgg cgacactagg gataacggaa

661 ctggagtatg agttttaccg tcggctggtg gaaaaacagt tgctcaccta tcgcctgcac

721 ggtgagtcgt ggcgtgaaaa agtgatcgaa cgtccggtgg tacataaaga ttacgatgaa

781 cagccgcgcg ggccgtttat tgtctgtgtg gatacttccg gctcaatggg cggctttaat

841 gaacagtgtg cgaaagcgtt ctgcctggcc ttgatgcgca ttgctctcgc agaaaaccgg

901 cgctgctata ttatgctatt ttccaccgag atcgtccgtt atgagctttc aggcccacaa

961 ggcatcgaac aagcaatccg ttttttaagc cagcagtttc gtggcggcac cgatcttgcc

1021 agttgttttc gcgccattat ggaacgcttg caaagcaggg aatggtttga tgccgatgcg

1081 gtggtgattt ctgattttat cgctcagcgg ttgcctgacg acgtgacgag taaagtgaaa

1141 gagctgcagc gggtacatca gcatcgcttt catgccgtgg cgatgtcggc acacggcaaa

1201 cccggcatca tgcgcatttt cgatcatatc tggcgctttg ataccgggat gcgaagccgc

1261 ctgctcagac gctggcggcg ataa

SEQ ID NO: 4 (Protein AAC76768.1; ADS2)

1 mrsrlkdarv ppelteevmc yqqsqllstp qfivqlpqil dllhrlnspw aeqarqlvda

61 nstitsalht lflqrwrlsl ivqattlnqq lleereqll sevqermtls gqlepiladn

121 ntaagrlwdm sagqlkrgdy qlivkygefl neqpelkrla eqlgrsreak siprndaqme

181 tfrtmvrepa tvpeqvdglq qsddilrllp pelatlgite leyefyrrlv ekqlltyrlh

241 geswrekvie rpvvhkdyde qprgpfivcv dtsgsmggfn eqcakafcla lmrialaenr

301 rcyimlfste ivryelsgpq gieqairfls qqfrggtdla scfraimerl qsrewfdada

361 vvisdfiaqr lpddvtskvk elqrvhqhrf havamsahgk pgimrifdhi wrfdtgmrsr

421 llrrwrr

SEQ ID NO: 5 (Nucleotide coding sequence for P10155 (ADS5) protein)

1 attttgcctt tttgttaggt ttcctaaaga caaaaaaaaa tggaggaatc tgtaaaccaa

61 atgcagccac tgaatgagaa gcagatagcc aattctcagg atggatatgt atggcaagtc

121 actgacatga atcgactaca ccggttctta tgtttcggtt ctgaaggtgg gacttattat

181 atcaaagaac agaagttggg ccttgaaaat gctgaagctt taattagatt gattgaagat

241 ggcagaggat gtgaagtgat acaagaaata aagtcattta gtcaagaagg cagaaccaca

301 aagcaagagc ctatgctctt tgcacttgcc atttgttccc agtgctccga cataagcaca

361 aaacaagcag catttaaagc tgtttctgaa gtttgtcgca ttcctaccca tctctttact

421 tttatccagt ttaagaaaga tctgaaggaa agcatgaaat gtggcatgtg gggtcgtgcc

481 ctccggaagg ctatagcgga ctggtacaat gagaaaggtg gcatggccct tgctctggca

541 gttacaaaat ataaacagag aaatggctgg tctcacaaag atctattaag attgtcacat

601 cttaaacctt ccagtgaagg acttgcaatt gtgaccaaat atattacaaa gggctggaaa

661 gaagttcatg aattgtataa agaaaaagca ctctctgtgg agactgaaaa attattaaag

721 tatctggagg ctgtagagaa agtgaagcgc acaaaagatg agctagaagt cattcatcta

781 atagaagaac atagattagt tagagaacat cttttaacaa atcacttaaa gtctaaagag

841 gtatggaagg ctttgttaca agaaatgccg cttactgcat tactaaggaa tctaggaaag

901 atgactgcta attcagtact tgaaccagga aattcagaag tatctttagt atgtgaaaaa

961 ctgtgtaatg aaaaactatt aaaaaaggct cgtatacatc catttcatat tttgatcgca

1021 ttagaaactt acaagacagg tcatggtctc agagggaaac tgaagtggcg ccctgatgaa

1081 gaaattttga aagcattgga tgctgctttt tataaaacat ttaagacagt tgaaccaact

1141 ggaaaacgtt tcttactagc tgttgatgtc agtgcttcta tgaaccaaag agttttgggt

1201 agtatactca acgctagtac agttgctgca gcaatgtgca tggttgtcac acgaacagaa

1261 aaagattctt atgtagttgc tttttccgat gaaatggtac catgtccagt gactacagat

1321 atgaccttac aacaggtttt aatggctatg agtcagatcc cagcaggtgg aactgattgc

1381 tctcttccaa tgatctgggc tcagaagaca aacacacctg ctgatgtctt cattgtattc

1441 actgataatg agacctttgc tggaggtgtc catcctgcta ttgctctgag ggagtatcga

1501 aagaaaatgg atattccagc taaattgatt gtttgtggaa tgacatcaaa tggtttcacc

1561 attgcagacc cagatgatag aggcatgttg gatatgtgcg gctttgatac tggagctctg

1621 gatgtaattc gaaatttcac attagatatg atttaaccat aagcagcagc acgatccaga

1681 gatccattgc catcagtgat ctcactaaaa aatatacagc tacttcccag ctaatctcca

1741 cccaatgaat gatgatggta tagtatgtgc ataatggaaa gttaccttac tg

SEQ ID NO: 6 (Protein P10155 (ADS5))

1 meesvnqmqp lnekqiansq dgyvwqvtdm nrlhrflcfg seggtyyike qklglenaea

61 lirliedgrg ceviqeiksf sqegrttkqe pmlfalaics qcsdistkqa afkavsevcr

121 ipthlftfiq fkkdlkesmk cgmwgralrk aiadwynekg gmalalavtk ykqrngwshk

181 dllrlshlkp sseglaivtk yitkgwkevh elykekalsv etekllkyle avekvkrtkd

241 elevihliee hrlvrehllt nhlkskevwk allqemplta llrnlgkmta nsvlepgnse

301 vslvceklcn ekllkkarih pfhilialet yktghglrgk lkwrpdeeil kaldaafykt

361 fktveptgkr fllavdvsas mnqrvlgsil nastvaaamc mvvtrtekds yvvafsdemv

421 pcpvttdmtl qqvlmamsqi paggtdcslp miwaqktntp advfivftdn etfaggvhpa

481 ialreyrkkm dipaklivcg mtsngftiad pddrgmldmc gfdtgaldvi rnftldmi

CLAIMS

1. A polypeptide, which polypeptide:

   (i) has the amino acid sequence as recited in SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6;

   (ii) is a fragment thereof having activity as an adhesion molecule or having an antigenic determinant in common with the polypeptide of (i); or

   (iii) is a functional equivalent of (i) or (ii).

2. A polypeptide which is a fragment according to claim 1(ii), which includes the adhesion molecule region of the ADS1 polypeptide, said adhesion molecule region being defined as including between residues 250 and 365 inclusive of the amino acid sequence recited in SEQ ID NO:2, wherein said fragment possesses the catalytic residues SER258, SER260 and ASP348, or equivalent residues, and possesses adhesion molecule activity.

3. A polypeptide which is a functional equivalent according to claim 1(iii), is homologous to the amino acid sequence as recited in SEQ ID NO:2, possesses the catalytic residues SER258, SER260 and ASP348, or equivalent residues, and has adhesion molecule activity.

4. A polypeptide according to claim 3, wherein said functional equivalent is homologous to the adhesion molecule region of the ADS1 polypeptide.

5. A polypeptide which is a fragment according to claim 1(ii), which includes the adhesion molecule region of the ADS2 polypeptide, said adhesion molecule region being defined as including between residue 267 and residue 384 of the amino acid sequence recited in SEQ ID NO:4, wherein said fragment possesses the catalytic residues SER273, SER275 and ASP365, or equivalent residues, and possesses adhesion molecule activity.

6. A polypeptide which is a functional equivalent according to claim 1(iii), is homologous to the amino acid sequence as recited in SEQ ID NO:4, possesses the

catalytic residues SER273, SER275 and ASP365, or equivalent residues, and has adhesion molecule activity.

7. A polypeptide according to claim 6, wherein said functional equivalent is homologous to the adhesion molecule region of the ADS2 polypeptide.

8. A polypeptide which is a fragment according to claim 1(ii), which includes the adhesion molecule region of the ADS5 polypeptide, said adhesion molecule region being defined as including between residue 373 and residue 503 of the amino acid sequence recited in SEQ ID NO:6, wherein said fragment possesses the catalytic residues SER378, SER380 and ASP469, or equivalent residues, and possesses adhesion molecule activity.

9. A polypeptide which is a functional equivalent according to claim 1(iii), is homologous to the amino acid sequence as recited in SEQ ID NO:6, possesses the catalytic residues SER378, SER380 and ASP469, or equivalent residues, and has adhesion molecule activity.

10. A polypeptide according to claim 9, wherein said functional equivalent is homologous to the adhesion molecule region of the ADS5 polypeptide.

11. A fragment or functional equivalent according to any one of claims 1-10, which has greater than 30% sequence identity with an amino acid sequence as recited in any one of SEQ ID NO:2, SEQ ID NO:4 and SEQ ID NO:6, or with a fragment thereof that possesses adhesion molecule activity, preferably greater than 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98% or 99% sequence identity, as determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for Biotechnology Information; http://www.ncbi.nlm.nih.gov/) [Blosum 62 matrix; gap open penalty=11 and gap extension penalty=1].

12. A functional equivalent according to any one of claims 1-10, which exhibits significant structural homology with a polypeptide having the amino acid sequence given in any one of SEQ ID NO:2, SEQ ID NO:4 and SEQ ID NO:6, or with a fragment thereof that possesses adhesion molecule activity.

13. A functional equivalent according to any one of claims 1-10, which exhibits significant structural homology with a polypeptide having the amino acid sequence given in any one of SEQ ID NO:2, SEQ ID NO:4 and SEQ ID NO:6, or with a fragment thereof that possesses adhesion molecule activity.

14. A fragment as recited in claim 1, 2, 5, 8, or 11, having an antigenic determinant in common with the polypeptide of claim 1(i), which consists of 7 or more (for example, 8, 10, 12, 14, 16, 18, 20 or more) amino acid residues from the sequence of SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6.

15. A purified nucleic acid molecule which encodes a polypeptide according to any one of the preceding claims.

16. A purified nucleic acid molecule according to claim 15, which has the nucleic acid sequence as recited in SEQ ID NO:1, SEQ ID NO:3 or SEQ ID NO:5, or is a redundant equivalent or fragment thereof.

17. A fragment of a purified nucleic acid molecule according to claim 15 or claim 16, which comprises between nucleotides 750 and 1095 of SEQ ID NO:1, or is a redundant equivalent thereof.

18. A fragment of a purified nucleic acid molecule according to claim 15 or claim 16, which comprises between nucleotides 801 and 1152 of SEQ ID NO:3, or is a redundant equivalent thereof.

19. A fragment of a purified nucleic acid molecule according to claim 15 or claim 16, which comprises between nucleotides 1119 and 1509 of SEQ ID NO:5, or is a redundant equivalent thereof.

20. A purified nucleic acid molecule which hybridizes under high stringency conditions with a nucleic acid molecule according to any one of claims 15-19.

21. A vector comprising a nucleic acid molecule as recited in any one of claims 15-20.

22. A host cell transformed with a vector according to claim 21.

23. A ligand which binds specifically to, and which preferably inhibits the adhesion molecule activity of, a polypeptide according to any one of claims 1-14.

24. A ligand according to claim 23, which is an antibody.

25. A compound that either increases or decreases the level of expression or activity of a polypeptide according to any one of claims 1-14.

26. A compound according to claim 25 that binds to a polypeptide according to any one of claims 1-14 without inducing any of the biological effects of the polypeptide.

27. A compound according to claim 25 or claim 26, which is a natural or modified substrate, ligand, enzyme, receptor or structural or functional mimetic.

28. A polypeptide according to any one of claim 1-14, a nucleic acid molecule according to any one of claims 15-20, a vector according to claim 21, a ligand according to claim 23 or 24, or a compound according to any one of claims 25-27, for use in therapy or diagnosis of disease.

29. A method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide according to any one of claim 1-14, or assessing the activity of a polypeptide according to any one of claim 1-14, in tissue from said patient and comparing said level of expression or activity to a control level, wherein a level that is different to said control level is indicative of disease.

30. A method according to claim 29 that is carried out *in vitro*.

31. A method according to claim 29 or claim 30, which comprises the steps of: (a) contacting a ligand according to claim 23 or claim 24 with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

32. A method according to claim 29 or claim 30, comprising the steps of:

a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule according to any one of claims 15-20 and the probe;

b) contacting a control sample with said probe under the same conditions used in step a); and

c) detecting the presence of hybrid complexes in said samples;

wherein detection of levels of the hybrid complex in the patient sample that differ from levels of the hybrid complex in the control sample is indicative of disease.

33. A method according to claim 29 or claim 30, comprising:

a) contacting a sample of nucleic acid from tissue of the patient with a nucleic acid primer under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule according to any one of claims 15-20 and the primer;

b) contacting a control sample with said primer under the same conditions used in step a); and

c) amplifying the sampled nucleic acid; and

d) detecting the level of amplified nucleic acid from both patient and control samples;

wherein detection of levels of the amplified nucleic acid in the patient sample that differ significantly from levels of the amplified nucleic acid in the control sample is indicative of disease.

34. A method according to claim 29 or claim 30 comprising:

a) obtaining a tissue sample from a patient being tested for disease;

b) isolating a nucleic acid molecule according to any one of claims 15-20 from said tissue sample; and

c) diagnosing the patient for disease by detecting the presence of a mutation which is associated with disease in the nucleic acid molecule as an indication of the disease.

35. The method of claim 34, further comprising amplifying the nucleic acid molecule to form an amplified product and detecting the presence or absence of a mutation in the amplified product.

36. The method of either claim 34 or 35, wherein the presence or absence of the mutation in the patient is detected by contacting said nucleic acid molecule with a nucleic acid probe that hybridises to said nucleic acid molecule under stringent conditions to form a hybrid double-stranded molecule, the hybrid double-stranded molecule having an

unhybridised portion of the nucleic acid probe strand at any portion corresponding to a mutation associated with disease; and

detecting the presence or absence of an unhybridised portion of the probe strand as an indication of the presence or absence of a disease-associated mutation.

37. A method according to any one of claims 29-36, wherein said disease is selected from cardiovascular diseases including atherosclerosis, ischaemia, restenosis, reperfusion injury, sepsis, haematological diseases such as leukaemia, blood clotting disorders, such as thrombosis, cancer including lung, prostate, breast, colorectal and brain tumours, metastasis, inflammatory diseases such as rhinitis, gastrointestinal diseases, including inflammatory bowel disease, ulcerative colitis, Crohn's disease, respiratory diseases including asthma, chronic obstructive pulmonary disease (COPD), respiratory distress syndrome, pulmonary fibrosis, immune disorders, including autoimmune diseases, rheumatoid arthritis, transplant rejection, allergy, liver diseases such as cirrhosis, endocrine diseases, such as diabetes, bone diseases such as osteoporosis, neurological diseases including stroke, multiple sclerosis, spinal cord injury, burns and wound healing, infections, preferably bacterial infection and most preferably *E. coli* infection.

38. Use of a polypeptide according to any one of claims 1-14 as an adhesion molecule.

39. Use of a nucleic acid molecule according to any one of claims 15-20 to express a protein that possesses adhesion molecule activity.

40. A method for effecting cell-cell adhesion, utilising a polypeptide according to any one of claims 1-14.

41. A pharmaceutical composition comprising a polypeptide according to any one of claims 1-14, a nucleic acid molecule according to any one of claims 15-20, a vector according to claim 21, a ligand according to claim 23 or 24, or a compound according to any one of claims 25-27.

42. A vaccine composition comprising a polypeptide according to any one of claims 1-14 or a nucleic acid molecule according to any one of claims 15-20.

43. A polypeptide according to any one of claims 1-14, a nucleic acid molecule according to any one of claims 15-20, a vector according to claim 21, a ligand according to claim 22 or 24, a compound according to any one of claims 25-27, or a pharmaceutical composition according to claim 41 for use in the manufacture of a medicament for the treatment of cardiovascular diseases including atherosclerosis, ischaemia, restenosis, reperfusion injury, sepsis, haematological diseases such as leukaemia, blood clotting disorders, such as thrombosis, cancer including lung, prostate, breast, colorectal and brain tumours, metastasis, inflammatory diseases such as rhinitis, gastrointestinal diseases, including inflammatory bowel disease, ulcerative colitis, Crohn's disease, respiratory diseases including asthma, chronic obstructive pulmonary disease (COPD), respiratory distress syndrome, pulmonary fibrosis, immune disorders, including autoimmune diseases, rheumatoid arthritis, transplant rejection, allergy, liver diseases such as cirrhosis, endocrine diseases, such as diabetes, bone diseases such as osteoporosis, neurological diseases including stroke, multiple sclerosis, spinal cord injury, burns and wound healing, infections, preferably bacterial infection and most preferably *E. coli* infection.

44. A method of treating a disease in a patient, comprising administering to the patient a polypeptide according to any one of claims 1-14, nucleic acid molecule according to any one of claims 15-20, a vector according to claim 21, a ligand according to claim 23 or 24, a compound according to any one of claims 25-27, or a pharmaceutical composition according to claim 41.

45. A method according to claim 44, wherein, for diseases in which the expression of the natural gene or the activity of the polypeptide is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an agonist.

46. A method according to claim 44, wherein, for diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide,

nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an antagonist.

47. A method of monitoring the therapeutic treatment of disease in a patient, comprising monitoring over a period of time the level of expression or activity of a polypeptide according to any one of claims 1-14, or the level of expression of a nucleic acid molecule according to any one of claims 15-20 in tissue from said patient, wherein altering said level of expression or activity over the period of time towards a control level is indicative of regression of said disease.

48. A method for the identification of a compound that is effective in the treatment and/or diagnosis of disease, comprising contacting a polypeptide according to any one of claims 1-14, a nucleic acid molecule according to any one of claims 15-20, or a host cell according to claim 21 with one or more compounds suspected of possessing binding affinity for said polypeptide or nucleic acid molecule, and selecting a compound that binds specifically to said nucleic acid molecule or polypeptide.

49. A kit useful for diagnosing disease comprising a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule according to any one of claims 15-20; a second container containing primers useful for amplifying said nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of disease.

50. The kit of claim 49, further comprising a third container holding an agent for digesting unhybridised RNA.

51. A kit comprising an array of nucleic acid molecules, at least one of which is a nucleic acid molecule according to any one of claims 15-20.

52. A kit comprising one or more antibodies that bind to a polypeptide as recited in any one of claims 1-14 and a reagent useful for the detection of a binding reaction between said antibody and said polypeptide.

53. A transgenic or knockout non-human animal that has been transformed to express higher, lower or absent levels of a polypeptide according to any one of claims 1-14.

54. A method for screening for a compound effective to treat disease, by contacting a non-human transgenic animal according to claim 53 with a candidate compound and determining the effect of the compound on the disease of the animal.

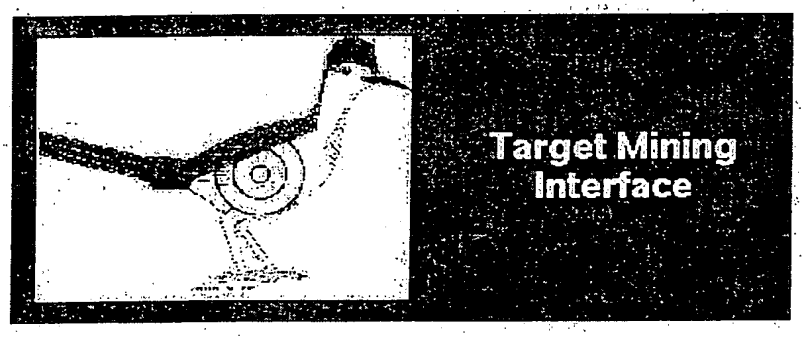
ABSTRACT

This invention relates to novel proteins, termed AAC74854.1, AAC76768.1 and P10155, herein identified as adhesion molecules and to the use of these proteins and nucleic acid sequences from the encoding genes in the diagnosis, prevention and treatment of disease.

Figure 1

File Edit View Go Communicator Help

Bookmarks Location http://victoria.inpharmatica.co.uk/~volker/BPD3target.html



inpharmatica

Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): 1lfa and chain (e.g. B): a

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release: DEVF9 - BPD3

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

Matches detected during the first 20 forward iterations

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

- Filter for the following SPECIES:

Homo sapiens | Rattus norvegicus (Rat) | Mus musculus (Mouse) | Danio rerio (Zebrafish)

100%

Figure 2A

File Edit View Go Communicator Help

Bookmarks Location: http://london-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl

2) 84 additional hits identified by both, Genome Threader and PSI-BLAST:

Combined Genome Threader and PSI - Blast output: PSI - BLAST values are shown in maroon!

Add2list	BPD links	WWW link	Title	Organism	Div.	%ID (GT,PSI)	Query rgn (GT,PSI)	Target rgn. (GT,PSI)	Aln. score (GT)	Conf. (GT)	1st. Iter. (PSI)	Best Iter. (PSI)	Best E-val (PS
☐	AAA59544.1 drill through Top50BlastHits Red.Seq.View	AAA59544.1	Not given	Homo sapiens	PRI	33.7%, 33% unmaskedSW	4-183, 4-183	150-336, 150-336	449	100% unmaskedGT	1	2	3E-
→ ☐	AAB24821.1 drill through Top50BlastHits Red.Seq.View	AAB24821.1	leukocyte integrin alpha chain	Homo sapiens	PRI	33.7%, 33% unmaskedSW	4-183, 4-183	150-336, 150-336	449	100% unmaskedGT	1	2	3E-
☐	Q99715 drill through Top50BlastHits Red.Seq.View	Q99715	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR.	Homo sapiens (Human).	PRI	28.8%, 26% unmaskedSW	4-181, 2-174	140-318, 2321-2495	440	100% unmaskedGT	1	3	2E-
☐	P20701 drill through Top50BlastHits Red.Seq.View	P20701	LEUKOCYTE ADHESION GLYCOPROTEIN LFA-1 ALPHA CHAIN PRECURSOR (LEUKOCYTE FUNCTION ASSOCIATED MOLECULE 1, ALPHA CHAIN) (CD11A) (INTEGRIN ALPHA-L).	Homo sapiens (Human).	PRI	100%, 100% unmaskedSW	1-183, 1-183	153-335, 153-335	423	100% unmaskedGT	1	1	1E-
☐	AAC31672.1 drill through Top50BlastHits Red.Seq.View	AAC31672.1	leukocyte function-associated molecule-1 alpha subunit	Homo sapiens	PRI	99.5%, 99% unmaskedSW	1-183, 1-183	153-335, 153-335	423	100% unmaskedGT	1	1	1E-
☐	CAA72402.1 drill through Top50BlastHits Red.Seq.View	CAA72402.1	collagen type XIV	Homo sapiens	PRI	29.1%, 29% unmaskedSW	2-180, 2-180	5-185, 5-185	422	100% unmaskedGT	1	2	2E-
☐	AAB38702.1 drill through Top50BlastHits Red.Seq.View	AAB38702.1	cartilage matrix protein	Homo sapiens	PRI	31.7%, 27% unmaskedSW	4-183, 2-182	275-455, 39-223	413	100% unmaskedGT	1	2	2E-
☐	CAB70853.1 drill through Top50BlastHits Red.Seq.View	CAB70853.1	hypothetical protein	Homo sapiens	PRI	28%, 28% unmaskedSW	1-180, 1-183	437-620, 437-624	406	100% unmaskedGT	1	2	1E-
☐	CAA27972.1 drill through Top50BlastHits Red.Seq.View	CAA27972.1	Not given	Homo sapiens	PRI	20.5%, 20% unmaskedSW	3-183, 2-181	1497-1673, 1689-1873	405	100% unmaskedGT	2	3	1E-
☐	AAB59512.1 drill through Top50BlastHits Red.Seq.View	AAB59512.1	Not given	Homo sapiens	PRI	20.5%, 20% unmaskedSW	3-183, 2-181	758-934, 950-1134	405	100% unmaskedGT	2	3	1E-
☐	CAA07569.1 drill through Top50BlastHits Red.Seq.View	CAA07569.1	matrilin-4	Homo sapiens	PRI	28.1%, 25% unmaskedSW	1-183, 1-183	342-528, 31-217	403	100% unmaskedGT	1	2	2E-

File Edit View Go Communicator Help

Bookmarks Location: http://london-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl

Fig 21

1) 509 hits identified by Genome Threader only:

Add2list	Redundant sequence display	BPD link	WWW link	Title	Organism	Div.	%ID	Query rgn.	Target rgn.	Aln. score	Conf.
☐	Red.Seq.View	AAF71133.1 drill through Top50BlastHits	AAF71133.1	PRO2769	Homo sapiens	PRI	13.8% unmaskedSW	109-179	1-80	122	100% unmaskedGT
☐	Red.Seq.View	CAB52192.1 drill through Top50BlastHits	CAB52192.1	G7c protein	Homo sapiens	PRI	9.6% unmaskedSW	7-117	20-124	82	99.18% unmaskedGT
☐	Red.Seq.View	CAA82910.1 drill through Top50BlastHits	CAA82910.1	basic transcripion factor 2, 44 kD subunit	Homo sapiens	PRI	11% unmaskedSW	5-164	61-225	75	99.18% unmaskedGT
☐	Red.Seq.View	AAD21820.1 drill through Top50BlastHits	AAD21820.1	NG37	Homo sapiens	PRI	9.6% unmaskedSW	7-117	318-422	82	98.86% unmaskedGT
☐	Red.Seq.View	BAA20761.1 drill through Top50BlastHits	BAA20761.1	Not given	Homo sapiens	PRI	16.4% unmaskedSW	5-114	1836-1950	78	98.52% unmaskedGT
☐	Red.Seq.View	AAA36154.1 drill through Top50BlastHits	AAA36154.1	Not given	Homo sapiens	PRI	13.2% unmaskedSW	5-112	10-137	79	98.16% unmaskedGT
☐	Red.Seq.View	AAF03046.1 drill through Top50BlastHits	AAF03046.1	candidate tumor suppressor protein DICE1	Homo sapiens	PRI	13.8% unmaskedSW	5-113	4-131	79	97.07% unmaskedGT
→	Red.Seq.View	AAC74854.1 drill through Top50BlastHits	AAC74854.1	orf, hypothetical protein	Escherichia coli	BCT	16.7% unmaskedSW	5-99	250-335	78	95.09% unmaskedGT
☐	Red.Seq.View	AAB60942.1 drill through Top50BlastHits	AAB60942.1	breast cancer suppressor candidate 1	Homo sapiens	PRI	16.2% unmaskedSW	63-156	2-90	80	93.33% unmaskedGT
☐	Red.Seq.View	AAA67537.1 drill through Top50BlastHits	AAA67537.1	glycoprotein IIIa	Homo sapiens	PRI	18.9% unmaskedSW	3-112	111-251	72	92.08% unmaskedGT reverse Hit
☐	Red.Seq.View	AAA52589.1 drill through Top50BlastHits	AAA52589.1	Not given	Homo sapiens	PRI	19.6% unmaskedSW	3-112	138-277	74	91.34% unmaskedGT reverse Hit
☐	Red.Seq.View	AAB71380.1 drill through Top50BlastHits	AAB71380.1	platelet membrane glycoprotein IIIa beta subunit	Homo sapiens	PRI	19.6% unmaskedSW	3-112	138-277	74	91.34% unmaskedGT reverse Hit
☐	Red.Seq.View	AAA35927.1 drill through Top50BlastHits	AAA35927.1	Not given	Homo sapiens	PRI	19.6% unmaskedSW	3-112	138-277	74	91.34% unmaskedGT reverse Hit

630 out of these 645 PSI-BLAST matches were identified using 'positive iterations':

P20701 ,1LFA:F ,1EDO:A ,1EDO:B ,1EOP:B ,1EOP:A ,1EDN ,AAC31672.1 ,1DGQ:A ,AAD25884.1 ,AAD25885.1 ,AAA39426.1 ,CAA30479.1 ,AAA59544.1 ,1BHO:4 ,1BHO:2 ,1BHO:3 ,1IDN2 ,1IDN4 ,1BHO:2 ,AAB24821.1 ,1ILM ,AAB16869.1 ,1IDO ,AAF21241.1 ,AAB38702.1 ,CAA06890.1 ,CAA06889.1 ,CAB70853.1 ,P05099 ,AAC53163.1 ,CAA07563.1 ,CAB46380.1 ,AAA39719.1 ,P51942 ,BAA00701.1 ,AAB38547.1 ,Q99715 ,CAA50063.1 ,P32018 ,Q99715 ,CAA36384.1 ,1CK4:B ,AAA39719.1 ,CAA72402.1 ,BAA00701.1 ,1CK4:A ,AAC01506.1 ,Q91145 ,P56199 ,CAB71222.1 ,BAA23160.1 ,1QC5:B ,CAA50063.1 ,P32018 ,AAA39719.1 ,Q99715 ,AAA82929.1 ,1QC5:A ,AAD04919.1 ,AAB93766.1 ,CAA68283.1 ,P17301 ,1AOX:A ,1AOX:B ,AAC52142.1 ,AAF43107.1 ,P38570 ,AAB59255.1 ,CAA27972.1 ,AAB59512.1 ,CAB37653.1 ,AAA51620.1 ,CAA12110.1 ,AAC23663.1 ,CAB86200.1 ,AAF01258.1 ,AAD51919.2 ,CAA82977.1 ,BAA00701.1 ,CAB37651.1 ,CAA03885.1 ,CAB37847.1 ,CAB86185.1 ,AAB51554.1 ,CAB86202.1 ,AAB51548.1 ,AAC31952.1 ,AAB65641.1 ,CAB37852.1 ,AAB51556.1 ,AAD50539.1 ,CAB37856.1 ,Q99715 ,AAB93226.1 ,CAB37658.1 ,CAB37666.1 ,CAB86188.1 ,CAB37672.1 ,BAA36034.1 ,AAF25351.1 ,CAB40275.1 ,CAB53844.1 ,CAB37665.1 ,AAB38798.1 ,AAD26126.1 ,AAB51557.1 ,CAB37670.1 ,CAA36267.1 ,AAF25353.1 ,AAB51550.1 ,CAB37667.1 ,AAB65641.1 ,AAC53163.1 ,AAB51547.1 ,P05099 ,CAA78125.1 ,AAB51546.1 ,AAB38702.1 ,CAA07563.1 ,AAF42925.1 ,AAB65642.1 ,AAD50544.1 ,AAB51552.1 ,AAD50541.1 ,AAB38799.1 ,AAF42927.1 ,AAD50540.1 ,CAB37672.1 ,AAA39719.1 ,AAC32844.1 ,AAB51549.1 ,CAB58678.1 ,AAB65642.1 ,CAA27972.1 ,AAB59512.1 ,1OAK:A ,1AUQ ,CAB37670.1 ,AAB51558.1 ,CAB65984.1 ,AAB38799.1 ,AAC32845.1 ,AAB51557.1 ,AAB51550.1 ,AAD26126.1 ,CAB37857.1 ,AAA20322.1 ,AAC72024.1 ,P51942 ,AAB65643.1 ,AAD26125.1 ,CAB37849.1 ,AAF23492.1 ,CAB46380.1 ,AAB51551.1 ,CAB37667.1 ,CAB37843.1 ,AAD50539.1 ,CAB37665.1 ,AAC32845.1 ,AAF25354.1 ,AAD25428.1 ,CAA06885.1 ,CAB37849.1 ,CAB37846.1 ,AAD04919.1 ,AAB93766.1 ,CAB86200.1 ,AAC62253.1 ,AAD01978.1 ,CAB37843.1 ,AAB51559.1 ,CAB37847.1 ,CAB86184.1 ,AAA03201.1 ,CAB37857.1 ,CAB86202.1 ,CAB37856.1 ,AAD50543.1 ,AAB51560.1 ,AAF25352.1 ,AAC32845.1 ,AAC32844.1 ,AAF25351.1 ,CAA36267.1 ,AAB51553.1 ,CAA97406.1 ,AAA03201.1 ,CAA71532.1 ,AAA03201.1 ,AAC39949.1 ,O43405 ,AAA03201.1 ,AAB51556.1 ,AAB51551.1 ,CAB37651.1 ,AAD01978.1 ,AAD40483.1 ,AAD01701.1 ,AAC32846.1 ,AAC62253.1 ,AAD01978.1 ,O43405 ,CAA36267.1 ,AAC23667.1 ,AAC39949.1 ,CAA36267.1 ,AAD01978.1 ,AAC23668.1 ,AAC06229.1 ,CAA36267.1 ,AAF19243.1 ,CAA36267.1 ,AAA03201.1 ,AAC23667.1 ,BAA23627.1 ,CAA36267.1 ,AAD01978.1 ,AAD01978.1 ,AAC23667.1 ,BAA00701.1 ,AAC68807.1 ,AAC23667.1 ,AAA03201.1 ,AAD01978.1 ,AAC23667.1 ,CAA36267.1 ,AAF19243.1 ,AAA58965.1 ,Q02388 ,AAB66593.1 ,AAC46572.1 ,AAC06229.1 ,CAB86227.1 ,AAF25353.1 ,AAD01701.1 ,CAB86203.1 ,AAD26126.1 ,AAF42927.1 ,CAB86184.1 ,AAD50540.1 ,AAF25354.1 ,AAD26125.1 ,AAB21220.2 ,CAB37653.1 ,AAB51553.1 ,AAB51549.1 ,AAD26127.1 ,CAB86185.1 ,AAB51555.1 ,AAA96311.1 ,CAB37852.1 ,AAB38797.1 ,AAD50542.1 ,AAB38800.1 ,CAB37668.1 ,CAB37846.1 ,AAB38802.1 ,AAF25352.1 ,AAB38798.1 ,CAB86203.1 ,AAC06229.1 ,CAB86188.1 ,AAB51554.1 ,CAB86227.1 ,AAB21220.2 ,AAB51545.1 ,AAF42925.1 ,AAA36734.1 ,AAF42926.1 ,AAB51546.1 ,AAB51552.1 ,AAD26124.1 ,AAB38802.1 ,AAD50544.1 ,AAB51549.1 ,AAB51545.1 ,AAB38797.1 ,CAB37666.1 ,AAB51555.1 ,AAA29773.1 ,AAA29774.1 ,AAA29776.1 ,BAA31171.1 ,BAA31182.1 ,BAA31188.1 ,BAA31193.1 ,BAA31174.1 ,BAA31178.1 ,AAA29772.1 ,CAA31440.1 ,BAA31169.1 ,AAC18657.1 ,BAA31172.1 ,BAA31175.1 ,BAA31176.1 ,BAA31195.1 ,BAA31194.1 ,AAB51558.1 ,BAA31187.1 ,AAA29771.1 ,AAA29778.1 ,AAA29767.1 ,AAA29775.1 ,AAA29770.1 ,BAA31183.1 ,BAA31190.1 ,BAA31167.1 ,BAA31177.1 ,AAD50543.1 ,BAA31170.1 ,CAA46541.1 ,CAA33891.1 ,AAA35619.1 ,AAA35619.1 ,AAF42926.1 ,AAA96311.1 ,AAA03201.1 ,CAA36267.1 ,CAA33144.1 ,CAA33981.1 ,AAB63303.1 ,AAA29076.1 ,AAD04919.1 ,AAB93766.1 ,CAA27972.1 ,AAB59512.1 ,1AO3:A ,1AO3:B ,1ATE:B ,1ATE:A ,AAA82929.1 ,AAB38801.1 ,CAA79152.1 ,AAD26127.1 ,CAA33889.1 ,CAA67576.1 ,AAD01978.1 ,AAD50542.1 ,AAC37484.1 ,AAC47463.1 ,AAF01565.1 ,Q04457 ,AAC23667.1 ,AAB59564.1 ,AAA03201.1 ,AAB34889.1 ,AAC47462.1 ,AAB63302.1 ,CAA71401.1 ,AAA03201.1 ,AAD03350.1 ,AAB97607.1 ,BAA78416.1 ,Q04457 ,CAA33888.1 ,AAB38800.1 ,CAB71222.1 ,AAC86162.1 ,CAA72303.1 ,AAA37381.1 ,AAA37380.1 ,CAA51389.1 ,BAA06179.1 ,CAA47744.1 ,AAC46961.1 ,BAA02763.1 ,BAA34707.1 ,AAC79582.1 ,CAB38978.1 ,CAA79153.1 ,CAA46541.1 ,AAF53476.1 ,AAF44957.1 ,BAA82322.1 ,AAC84160.1 ,BAA34706.1 ,CAB03055.1 ,CAB03054.1 ,AAC47957.1 ,AAB51547.1 ,AAC47957.1 ,CAA98248.1 ,AAB59564.1 ,CAB46421.1 ,CAA33891.1 ,AAF00021.1 ,CAA79153.1 ,CAA67576.1 ,AAA50715.1 ,AAF13350.1 ,AAC24429.1 ,AAB24637.1 ,CAB03057.1 ,AAB52325.1 ,AAB59564.1 ,Q04457 ,BAA02853.1 ,CAB38978.1 ,AAC46961.1 ,AAC47461.1 ,CAB75310.1 ,CAA79152.1 ,CAB38978.1 ,AAC46961.1 ,AAD16289.1 ,CAB02952.1 ,BAA82322.1 ,AAC23667.1 ,CAA79156.1 ,CAA33889.1 ,AAD30859.1 ,AAC23667.1 ,CAA79156.1 ,AAB38801.1 ,BAA82322.1 ,CAA33144.1 ,BAA23627.1 ,CAA36267.1 ,AAA03201.1 ,CAA33144.1 ,CAA33981.1 ,AAD01978.1 ,AAC19251.1 ,AAD50541.1 ,AAA03201.1 ,CAA31487.1 ,BAA82322.1 ,AAB04590.2 ,CAA61109.1 ,AAD50541.1 ,AAB51560.1 ,AAC17639.1 ,AAC46572.1 ,AAB51559.1 ,CAB38978.1 ,AAA36795.1 ,AAC46961.1 ,AAA36795.1 ,CAA36267.1 ,AAA36795.1 ,AAF19243.1 ,AAD29428.1 ,BAA07536.1 ,AAF69610.1 ,AAC36289.1 ,AAA81562.1 ,AAF03259.1 ,AAB50139.1 ,AAB50141.1 ,P54290 ,AAB50142.1 ,AAB50140.1 ,AAB50138.1 ,Q61703 ,BAA13939.1 ,CAA72308.1 ,P19823 ,P54289 ,CAA72155.1 ,Q61704 ,CAB86192.1 ,AAC70914.1 ,CAB86193.1 ,Q63416 ,CAA30160.1 ,BAA13940.1 ,CAA05423.1 ,CAA49841.1 ,CAB75878.1 ,CAB75362.1 ,BAA07632.1 ,AAB46821.1 ,BAA13938.1 ,BAA88322.1 ,AAD48398.1 ,CAB77598.1 ,P76481 ,AAC79610.1 ,Q55874 ,AAF58395.1 ,CAB03059.1 ,AAF53505.1 ,AAF44988.1 ,CAB03056.1 ,CAB03058.1 ,CAA88984.1 ,AAC25786.1 ,AAA36225.2 ,CAA88985.1 ,BAA12207.1 ,BAA16787.1 ,CAA88986.1 ,O50313 ,AAD15576.1 ,AAF70042.1 ,CAB55466.1 ,AAB63940.1 ,AAC79447.1 ,BAA08371.1 ,CAA33890.1 ,P34392 ,AAF63021.1 ,AAB19093.1 ,AAF10447.1 ,AAD30858.1 ,AAC65716.1 ,CAA46541.1 ,AAC72024.1 ,AAC62540.1 ,AAF12450.1 ,AAD41583.1 ,BAA18152.1 ,CAB80454.1 ,AAC05096.1 ,AAC65016.1 ,AAB63363.1 ,AAB52325.1 ,AAF27914.1 ,CAB38978.1 ,AAC46961.1 ,CAA67559.1 ,BAA82322.1 ,CAB38978.1 ,AAC46961.1 ,CAA45188.1 ,CAA49279.1 ,AAC79982.1 ,AAC35003.1 ,AAC44511.1 ,AAF12731.1 ,AAB86529.1 ,AAB86530.1 ,AAB86531.1 ,CAB10840.1 ,AAD52031.1 ,CAB58179.1 ,AAC95429.1 ,Q22437 ,BAA33743.1 ,CAB63268.1 ,AAC62491.1 ,AAD47250.1 ,AAF22895.1 ,AAB38810.1 ,BAA00016.1 ,CAA88388.1 ,AAD42198.2 ,BAA20346.1 ,AAF70056.1 ,AAD51108.1 ,AAF27140.1 ,AAD31491.1 ,AAF23117.1 ,CAA80121.1 ,AAD31445.2 ,CAA16388.1 ,CAB03680.1 ,AAA27963.1 ,CAA47439.1 ,CAA77537.1 ,CAB08456.1 ,CAA90031.1 ,CAB71206.1 ,AAC75134.1 ,CAB03681.1 ,AAC84032.1 ,O34845 ,AAB85061.1 ,BAA76532.1 ,AAC83659.1 ,AAA68973.1 ,CAA10812.1 ,CAA91030.1 ,AAA92391.1 ,AAA60558.1 ,AAC44737.1 ,AAC45374.1 ,BAA10368.1 ,AAC83698.1 ,AAF39956.1 ,AAC75182.1 ,AAC46961.1 ,BAA91707.1 ,AAD25169.1 ,Q60384 ,CAB44273.1 ,AAD31497.2 ,AAD42206.2 ,AAD51756.1 ,AAF57485.1 ,CAB76889.1 ,AAB88815.1 ,AAC76768.1 ,BAA79575.1 ,AAD40367.1 ,AAF00077.1 ,CAB52078.1 ,AAB64957.1 ,AAF11936.1 ,BAA61542.1 ,BAA07331.1 ,CAA16691.1 ,CAA45619.1 ,BAA25430.1 ,AAB98912.1 ,AAD31504.2 ,CAB13830.1 ,CAB01991.1 ,AAD31490.2 ,AAF62872.1 ,CAB11610.1 ,AAA86852.2 ,CAB43000.1 ,AAD31503.2 ,AAB01767.1 ,AAD42203.3 ,AAD26900.1 ,AAD32681.3 ,CAB76247.1 ,AAD42207.2 ,CAA83139.2 ,P34567 ,CAA58282.1 ,AAA85219.1 ,

15 out of these 645 PSI-BLAST matches were identified using 'negative iterations':

BAA81233.1 ,CAA12066.1 ,CAB50193.1 ,CAA17374.1 ,AAB86281.1 ,AAB87741.1 ,AAA93386.1 ,AAA93386.1 ,AAA81430.1 ,AAC16991.1 ,BAA85551.1 ,AAB41649.1 ,AAC51223.1 ,AAC52946.1 ,Q01443 ,

A2) Genome Threader Matches:

2729 matches found by Genome Threading:

CAA30479.1 ,AAA59544.1 ,AAB24821.1 ,Q99715 ,AAA39719.1 ,AAB16869.1 ,CAA50063.1 ,P32018 ,BAA00701.1 ,P20701 ,AAC31672.1 ,P05099 ,CAA72402.1 ,AAB38702.1 ,P51942 ,AAD04919.1 ,AAB93766.1 ,AAD25884.1 ,AAA39426.1 ,AAD25885.1 ,CAA06890.1 ,CAA06889.1 ,CAB70853.1 ,CAA27972.1 ,AAB59512.1 ,CAA07563.1 ,CAB46380.1 ,AAC06229.1 ,CAB37653.1 ,AAC53163.1 ,AAB38547.1 ,CAB86185.1 ,CAB71222.1 ,AAF21241.1 ,AAB51544.1 ,CAB37651.1 ,CAB37672.1 ,AAA82929.1 ,AAC01506.1 ,Q91145 ,AAB65641.1 ,CAB37667.1 ,CAB37856.1 ,AAF25353.1 ,CAB37665.1 ,CAB86203.1 ,CAB86227.1 ,AAB51550.1 ,AAB51556.1 ,AAD01701.1 ,AAB51554.1 ,CAB37666.1 ,CAA36384.1 ,CAB37849.1 ,AAB38798.1 ,CAB86184.1 ,CAB37847.1 ,CAB37668.1 ,CAB65984.1 ,AAB51557.1 ,P56199 ,AAF25351.1 ,CAB86202.1 ,CAB37843.1 ,AAD50544.1 ,AAB38799.1 ,AAB51546.1 ,CAB86200.1 ,CAB37857.1 ,AAC32845.1 ,CAB37852.1 ,AAB21220.2 ,CAB86188.1 ,AAD50539.1 ,CAB37846.1 ,CAB37670.1 ,AAC32843.1 ,AAC62253.1 ,AAF25354.1 ,CAA68283.1 ,CAA03885.1 ,BAA36034.1 ,BAA23160.1 ,AAC32846.1 ,AAD26125.1 ,AAF42927.1 ,AAD50540.1 ,AAC32844.1 ,AAF01258.1 ,AAD51919.2 ,AAC52142.1 ,AAD26126.1 ,CAB37854.1 ,AAC23663.1 ,AAC62253.1 ,AAF25354.1 ,CAA68283.1 ,CAA03885.1 ,BAA36034.1 ,BAA23160.1 ,AAC32846.1 ,AAD26125.1 ,AAF42927.1 ,AAD50540.1 ,AAC32844.1 ,AAF01258.1 ,AAD51919.2 ,AAB65642.1 ,AAF43107.1 ,P38570 ,AAB59255.1 ,AAD26124.1 ,AAA51620.1 ,O43405 ,AAA29774.1 ,AAA29776.1 ,AAB51553.1 ,AAA29773.1 ,AAD50542.1 ,AAB38800.1 ,AAC39949.1 ,BAA31171.1 ,BAA31188.1 ,BAA31182.1 ,AAB65643.1 ,BAA31193.1 ,BAA31174.1 ,BAA31195.1 ,BAA31178.1 ,BAA31176.1 ,BAA31194.1 ,AAO18657.1 ,BAA31169.1 ,BAA31172.1 ,AAA29772.1 ,CAA31440.1 ,BAA31187.1 ,AAA29771.1 ,BAA31183.1 ,AAA29778.1 ,BAA31190.1 ,AAA29775.1 ,AAB03226.1 ,CAA36267.1 ,P17301 ,CAB40275.1 ,BAA31170.1 ,CAA82877.1 ,AAB51545.1 ,AAF25352.1 ,AAF19243.1 ,CAA46541.1 ,CAA33891.1 ,AAA35619.1 ,CAA12110.1 ,AAA29770.1 ,BAA31167.1 ,AAA29767.1 ,AAA29775.1 ,AAB03226.1 ,CAA36267.1 ,P17301 ,CAB40275.1 ,BAA31170.1 ,CAA82877.1 ,AAB51545.1 ,AAF25352.1 ,AAF19243.1 ,CAA46541.1 ,CAA33891.1 ,AAA35619.1 ,AAA35618.1 ,AAC31952.1 ,AAD01978.1 ,BAA82322.1 ,AAA29076.1 ,AAB51552.1 ,AAB38802.1 ,AAA36734.1 ,AAA03201.1 ,CAA78125.1 ,AAF23492.1 ,AAA20892.2 ,AAB51551.1 ,CAA33144.1 ,CAA33981.1 ,AAB38801.1 ,AAB38797.1 ,AAB51547.1 ,AAA35618.1 ,AAC31952.1 ,AAD01978.1 ,BAA82322.1 ,AAA29076.1 ,AAB51552.1 ,AAB38802.1 ,AAA36734.1 ,AAA03201.1 ,CAA78125.1 ,AAF23492.1 ,AAA20892.2 ,AAB51551.1 ,CAA33144.1 ,CAA33981.1 ,AAB38801.1 ,AAB38797.1 ,AAB51547.1 ,AAB63303.1 ,AAB51543.1 ,AAB66593.1 ,CAA97406.1 ,AAD50541.1 ,AAF01565.1 ,AAB51555.1 ,AAC72024.1 ,BAA23627.1 ,CAA79152.1 ,AAA96311.1 ,AAB51558.1 ,CAA71532.1 ,CAB58678.1 ,AAD26127.1 ,AAB86281.1 ,AAC37484.1 ,AAB63303.1 ,AAB51543.1 ,AAB66593.1 ,CAA97406.1 ,AAD50541.1 ,AAF01565.1 ,AAB51555.1 ,AAC72024.1 ,BAA23627.1 ,CAA79152.1 ,AAD29428.1 ,AAA96311.1 ,AAB51558.1 ,CAA71532.1 ,CAB58678.1 ,AAD26127.1 ,AAB86281.1 ,AAC37484.1 ,AAC47463.1 ,AAB51559.1 ,AAD50543.1 ,AAB51560.1 ,Q04457 ,BAA07536.1 ,AAF69610.1 ,Q61704 ,AAB87741.1 ,AAB63302.1 ,AAB59564.1 ,AAC47462.1 ,BAA13940.1 ,CAA33889.1 ,CAA67576.1 ,CAA72309.1 ,Q63416 ,AAC23668.1 ,AAA93386.1 ,AAC47463.1 ,AAB51559.1 ,AAD50543.1 ,AAB51560.1 ,Q04457 ,BAA07536.1 ,AAF69610.1 ,Q61704 ,AAB87741.1 ,AAB63302.1 ,AAB59564.1 ,AAC47462.1 ,BAA13940.1 ,CAA33889.1 ,CAA67576.1 ,CAA72309.1 ,Q63416 ,AAC23668.1 ,AAA93386.1 ,

Fig 3 Electronic

Inpharmatica Biopendium 1.4

File Edit Search Type Help

Redundant Sequence Display

Contains:
2 Sub-sequences
0 PROSITE hits
0 PRINTS hits

inpharmatica

Tools

Sequences

Ligands

Representative code: BAA15585.1 **Length:** 427 **Organism name:** *Escherichia coli* [...additional annotation]

Aligned sequences	Code
	BAA15585.1
	P76235
	AAC74854.1

Primary database information

GenBank

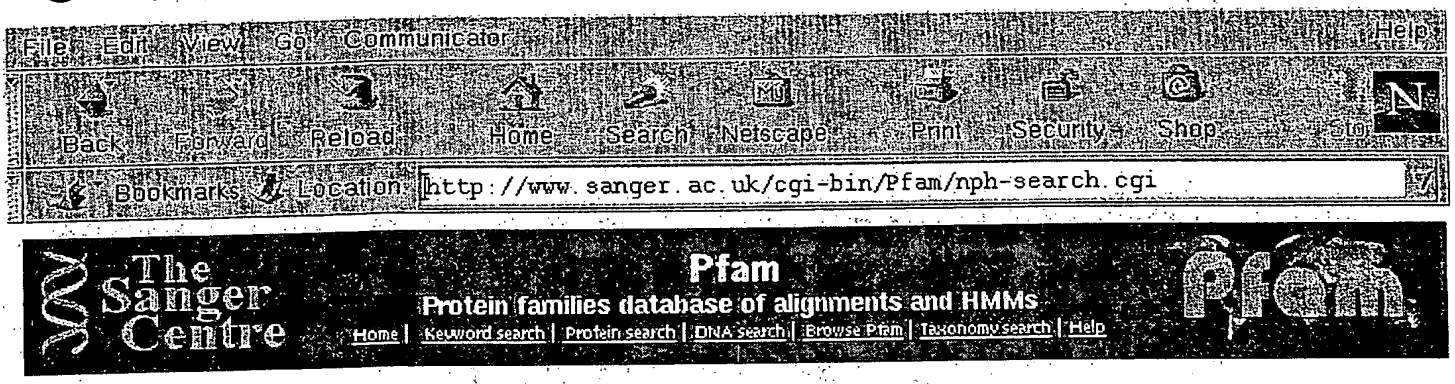
protein_id	Details	
BAA15585.1	**Tax ID:** 562	**Organism name:** *Escherichia coli*
	EC Number: *Not available*	**Gene Name:** yzdC
	Description: Not given	
	Other links: GI: 1736412 TAXON: 562	

SWISS-PROT

Code	Details	
P76235	**Tax ID:** 562	**Organism name:** *Escherichia coli*
	EC Number: *Not available*	**Gene Name:** YEAH
	Description: HYPOTHETICAL 49.4 KDA PROTEIN IN GAPA-RND INTERGENIC REGION	

Start | BPD1_FE1_4 | Inpharmati... | Fagan, Richa... | Inbox - Outlo... | Target Minin... | Microsoft Po...

Figure 4



Results for gi|1788084|gb|AAC74854.1|

There were no matches to Pfam-A (including borderline matches) for gi|1788084|gb|AAC74854.1|

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_39416	233	423	3.7e-103	Align

[427 residues]

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families: Hypertext linked to swisspfam

Query gi|1788084|gb|AAC74854.1|/233-423 matching Pfam-B_39416

```
                YEAH_ECOLI 233 DLRYKNYEKRPDPSSQAVMFCLMDVSGSMDQSTKDMAKRFYILLYLFLSR 282
                           DLRYKNYEKRPDPSSQAVMFCLMDVSGSMDQSTKDMAKRFYILLYLFLSR
gi|1788084|gb|AAC74854.1| 233 DLRYKNYEKRPDPSSQAVMFCLMDVSGSMDQSTKDMAKRFYILLYLFLSR 282

                YEAH_ECOLI 283 TYKNVEVVYIRHHTQAKEVDEHEFFYSQETGGTIVSSALKLMDEVVKERY 332
                           TYKNVEVVYIRHHTQAKEVDEHEFFYSQETGGTIVSSALKLMDEVVKERY
gi|1788084|gb|AAC74854.1| 283 TYKNVEVVYIRHHTQAKEVDEHEFFYSQETGGTIVSSALKLMDEVVKERY 332

                YEAH_ECOLI 333 NPAQWNIYAAQASDGDNWADDSPLCHEILAKKLLPVVRVYSYIEITRRAH 382
                           NPAQWNIYAAQASDGDNWADDSPLCHEILAKKLLPVVRVYSYIEITRRAH
gi|1788084|gb|AAC74854.1| 333 NPAQWNIYAAQASDGDNWADDSPLCHEILAKKLLPVVRVYSYIEITRRAH 382

                YEAH_ECOLI 383 QTLWREYEHLQSTFDNFAMQHIRDQDDIYPVFRELFHKQNA 423
                           QTLWREYEHLQSTFDNFAMQHIRDQDDIYPVFRELFHKQNA
gi|1788084|gb|AAC74854.1| 383 QTLWREYEHLQSTFDNFAMQHIRDQDDIYPVFRELFHKQNA 423
```

Align to family

If you think there is anything wrong with this script, please contact Pfam

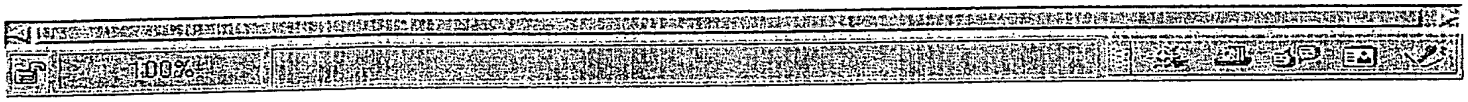


Figure 5

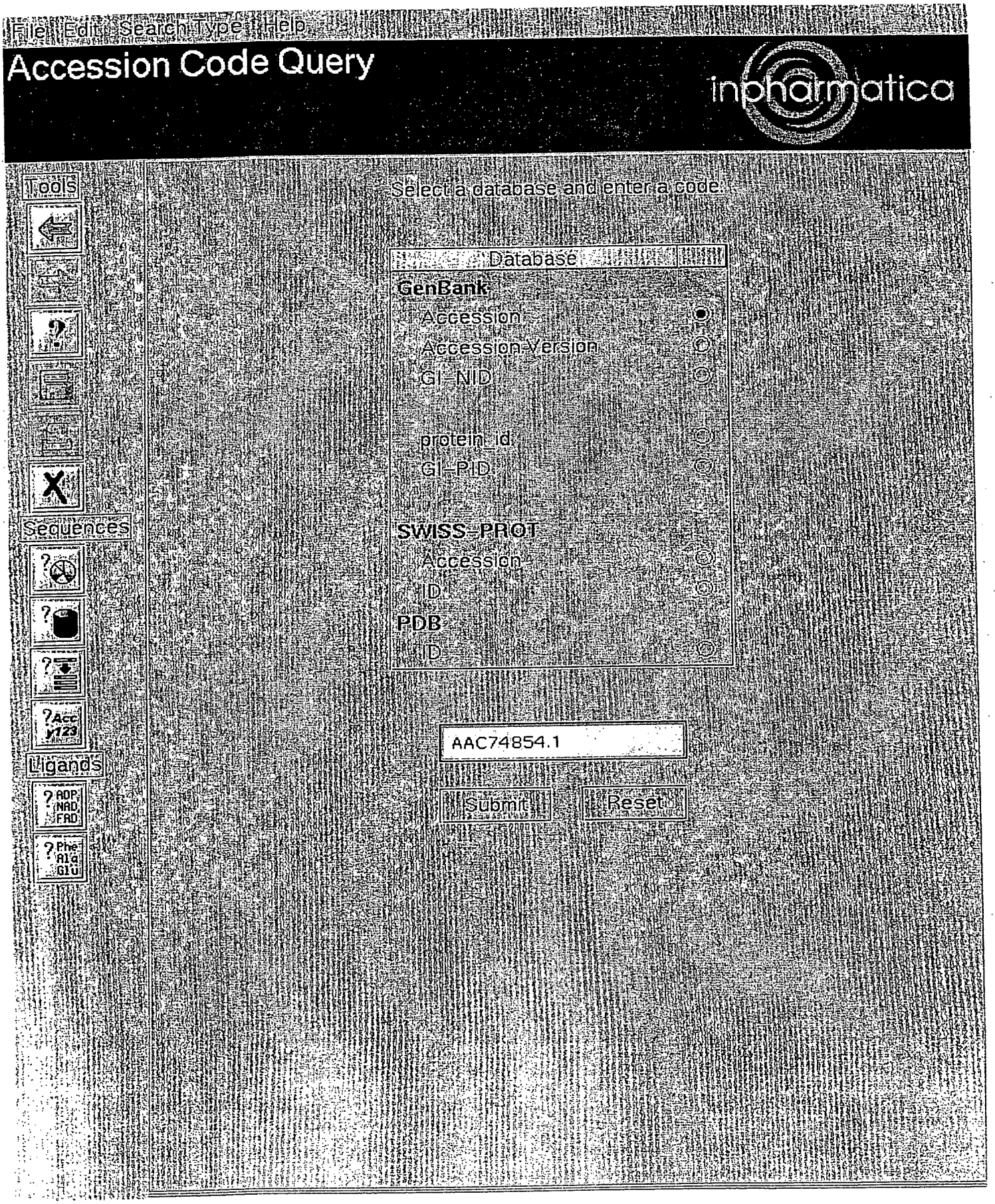
File Edit View Go Communicator Help

Bookmarks Location: http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=

```
LOCUS       AAC74854       427 aa                    BCT     01-DEC-2000
DEFINITION  orf, hypothetical protein [Escherichia coli K12].
ACCESSION   AAC74854
PID         g1788084
VERSION     AAC74854.1  GI:1788084
DBSOURCE    locus AE000273 accession AE000273.1
KEYWORDS    .
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (residues 1 to 427)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (residues 1 to 427)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (residues 1 to 427)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (residues 1 to 427)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..427
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     Protein         1..427
                     /function="orf; Unknown"
                     /product="orf, hypothetical protein"
     CDS             1..427
                     /gene="yeaH"
                     /coded_by="1788078:6385..7668"
                     /transl_table=11
                     /note="o427; This 427 aa ORF is 28 pct identical (43 gaps)
                     to 327 residues of an approx. 312 aa protein YZDC_BACSU
                     SW: P45742"
ORIGIN
        1 mtwfidrrln gknksmvnrq rflrrykaqi kqsiseaink rsvtdvdsge svsiptedis
       61 epmfhqgrgg lrhrvhpgnd hfvqndrier pqgggggsgs gqgqasqdge gqdefvfqis
      121 kdeyldllfe dlalpnlkqn qqrqltepkt hragytangv panisvvrsl qmslarrtam
      181 tagkrrelha leenlaiisn sepaqlieee rlrkeiaelr akiervpfid tfdlrykmye
```

100%

Figure 6A


File Edit Search Type Help
Accession Code Query
inpharmatica
Tools
Sequences
Ligands
?ADP NAD FAD
?Phe Ala Glu
?Acc y123
Select a database and enter a code:
Database
GenBank
Accession
Accession.Version
GI_NID
protein_id
GI_PID
SWISS_PROT
Accession
ID
PDB
ID
AAC74854.1
Submit
Reset

File Edit Search Type Help

Aligned Sequence Display

Query details:
AAC74854.1 orf, hypothetical protein

inpharmatica

Tools

Sequences

Ligands

Total selected: 0 | View alignment

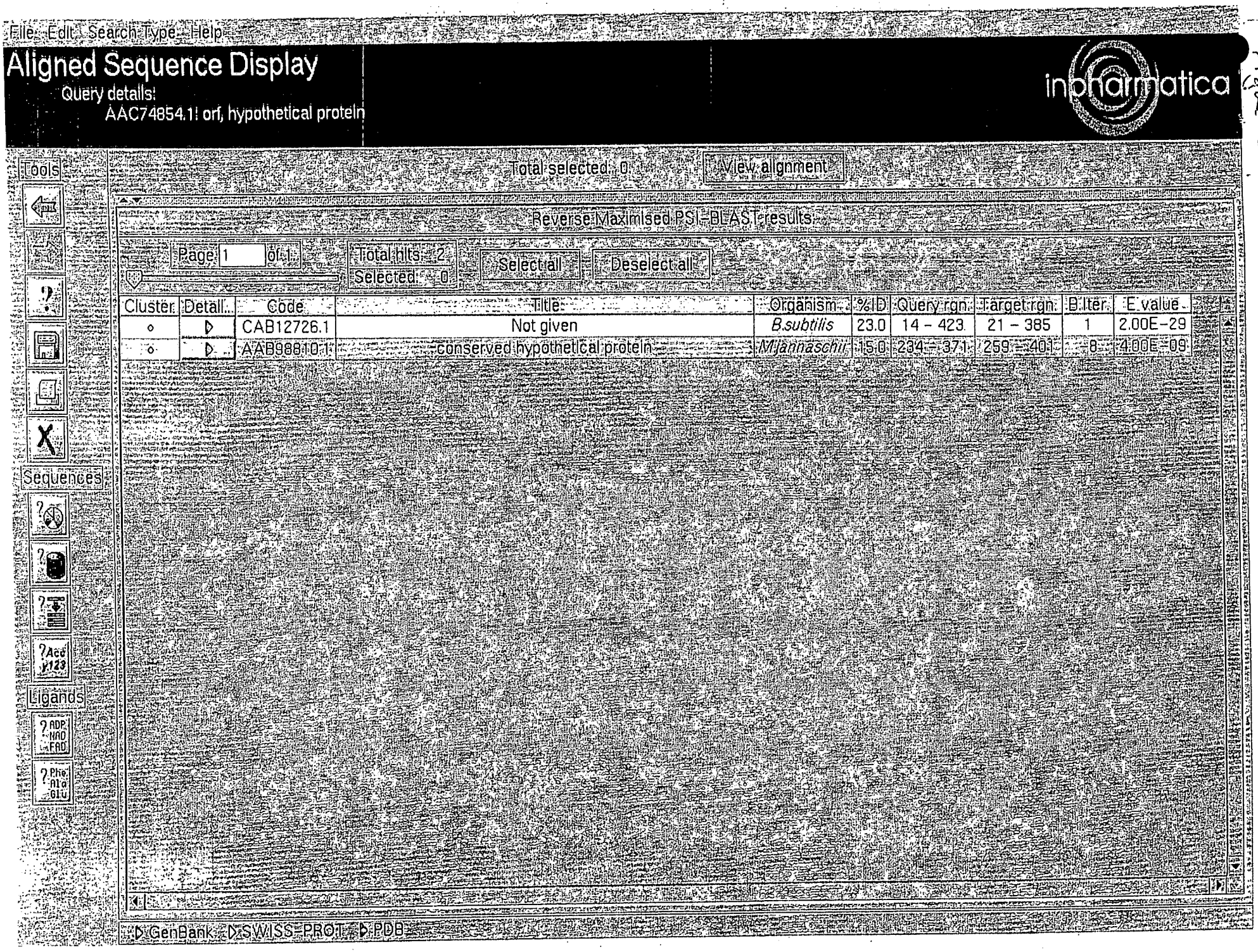
Inpharmatica Genome Threader results

Page 1 of 2 | Total hits: 36 | Selected: 0 | Select all | Deselect all

Cluster	Detail	Code	Title	Organism	%ID	Query rgn	Target rgn	Aln score	Method	Confidence
	→	1LFA:A	CD11A I-DOMAIN WITH BOUND MN++	*H.sapiens*	16.7	250 – 335	5 – 99	78	Local	Marginal(95%)
	▷	1ZOO:A	CD11A I-DOMAIN WITH BOUND MAGNESI...	*H.sapiens*	16.7	250 – 335	5 – 99	78	Local	Marginal(95%)
	▷	1ZOO:B	CD11A I-DOMAIN WITH BOUND MAGNESI...	*H.sapiens*	16.7	250 – 335	5 – 99	78	Local	Marginal(95%)
	▷	1ZOP:B	CD11A I-DOMAIN WITH BOUND MAGNESI...	*H.sapiens*	16.7	250 – 335	5 – 99	78	Local	Marginal(95%)
	▷	1ZOP:A	CD11A I-DOMAIN WITH BOUND MAGNESI...	*H.sapiens*	16.7	250 – 335	5 – 99	78	Local	Marginal(95%)
	▷	1BHO:1	MAC-1 I DOMAIN MAGNESIUM COMPLEX	*H.sapiens*	15.1	250 – 365	4 – 128	76	Local	Low(94%)
	▷	1IDN:1	MAC-1 I DOMAIN METAL FREE	*H.sapiens*	15.1	250 – 365	4 – 128	76	Local	Low(94%)
	▷	1BHQ:2	MAC-1 I DOMAIN CADMIUM COMPLEX	*H.sapiens*	15.1	250 – 365	4 – 128	76	Local	Low(94%)
	▷	1BHQ:1	MAC-1 I DOMAIN CADMIUM COMPLEX	*H.sapiens*	15.1	250 – 365	4 – 128	76	Local	Low(94%)
	▷	1BHO:2	MAC-1 I DOMAIN MAGNESIUM COMPLEX	*H.sapiens*	15.1	250 – 365	4 – 128	76	Local	Low(94%)
	▷	1IDN:2	MAC-1 I DOMAIN METAL FREE	*H.sapiens*	15.1	250 – 365	4 – 128	76	Local	Low(94%)
	▷	1ATZ:B	HUMAN VON WILLEBRAND FACTOR A3 DO...	*H.sapiens*	15.1	244 – 352	1 – 126	70	Local	Low(94%)
	▷	1ATZ:A	HUMAN VON WILLEBRAND FACTOR A3 DO...	*H.sapiens*	15.3	250 – 335	5 – 101	72	Local	Low(93%)
◇	▷	1DGQ:A	NMR SOLUTION STRUCTURE OF THE INSER...	*H.sapiens*	16.3	250 – 331	9 – 99	75	Local	Low(93%)
	▷	1IDO	I-DOMAIN FROM INTEGRIN CR3, MG2+ BO...	*H.sapiens*	15.1	250 – 365	4 – 128	75	Local	Low(92%)
	▷	1JLM	I-DOMAIN FROM INTEGRIN CR3, MN2+ BO...	*H.sapiens*	15.1	250 – 365	4 – 128	75	Local	Low(89%)
	▷	1ZON	CD11A I-DOMAIN WITHOUT BOUND CATI...	*H.sapiens*	16.7	250 – 335	5 – 99	72	Local	Low(89%)
	▷	1LFA:B	CD11A I-DOMAIN WITH BOUND MN++	*H.sapiens*	16.0	250 – 333	5 – 97	66	Local	Low(80%)
	▷	1OAK:A	CRYSTAL STRUCTURE OF THE VON WILLE...	*H.sapiens*	20.3	251 – 301	1 – 59	72	Local	Low(72%)
	▷	1AUQ	A1 DOMAIN OF VON WILLEBRAND FACTOR	*H.sapiens*	21.2	250 – 301	18 – 68	69	Local	Low(66%)
	▷	1AO3:A	A3 DOMAIN OF VON WILLEBRAND FACTOR	*H.sapiens*	15.2	245 – 331	2 – 99	63	Local	Low(53%)
	▷	1AO3:B	A3 DOMAIN OF VON WILLEBRAND FACTOR	*H.sapiens*	15.2	245 – 331	2 – 99	63	Local	Low(53%)
	▷	1AOX:B	I DOMAIN FROM INTEGRIN ALPHA2-BETA1	*H.sapiens*	13.1	242 – 335	1 – 105	53	Local	Low(21%)
◇	▷	1RYP:D	CRYSTAL STRUCTURE OF THE 20S PROTE...	*S.cerevisiae*	13.1	246 – 305	166 – 226	58	Local	Low(21%)
	▷	1RYP:R	CRYSTAL STRUCTURE OF THE 20S PROTE...	*S.cerevisiae*	13.1	246 – 305	166 – 226	58	Local	Low(21%)
	▷	1CK4:B	CRYSTAL STRUCTURE OF RAT A1B1 INTEG...	*R.norvegicus*	10.5	250 – 331	5 – 96	55	Local	Low(20%)

▷ GenBank ▷ SWISS-PROT ▷ PDB

Fig. 69


File Edit Search Type Help
Aligned Sequence Display
Query details:
AAC74854.1: orf, hypothetical protein
inpharmatica
Tools
Total selected: 0
View alignment
Reverse Maximised PSI-BLAST results
Page: 1 of 1
Total hits: 2
Selected: 0
Select all
Deselect all
Cluster | Detail | Code | Title | Organism | %ID | Query rgn | Target rgn | Iter | E value
CAB12726.1 | Not given | B.subtilis | 23.0 | 14 - 423 | 21 - 385 | 1 | 2.00E-29
AAB98810.1 | conserved hypothetical protein | M.jannaschii | 15.0 | 284 - 371 | 259 - 401 | 8 | 4.00E-09
Sequences
Ligands
GenBank
SWISS-PROT
PDB

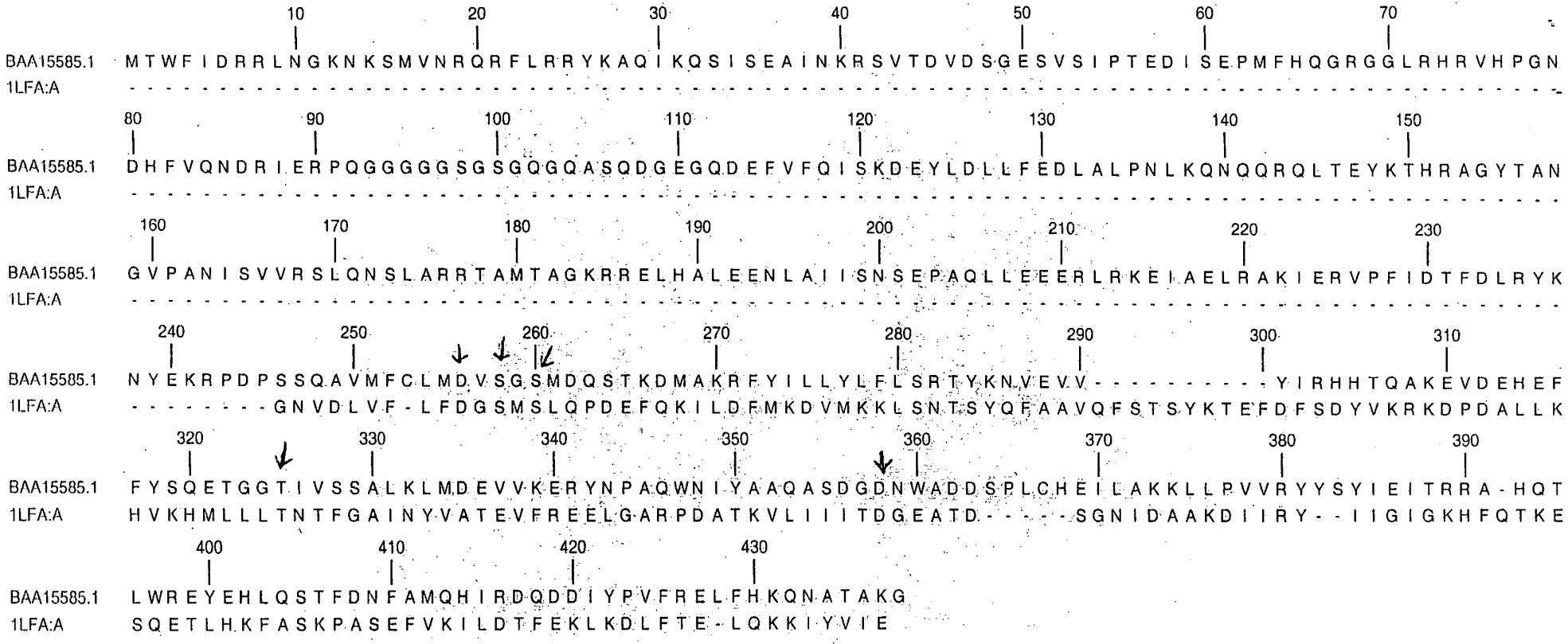
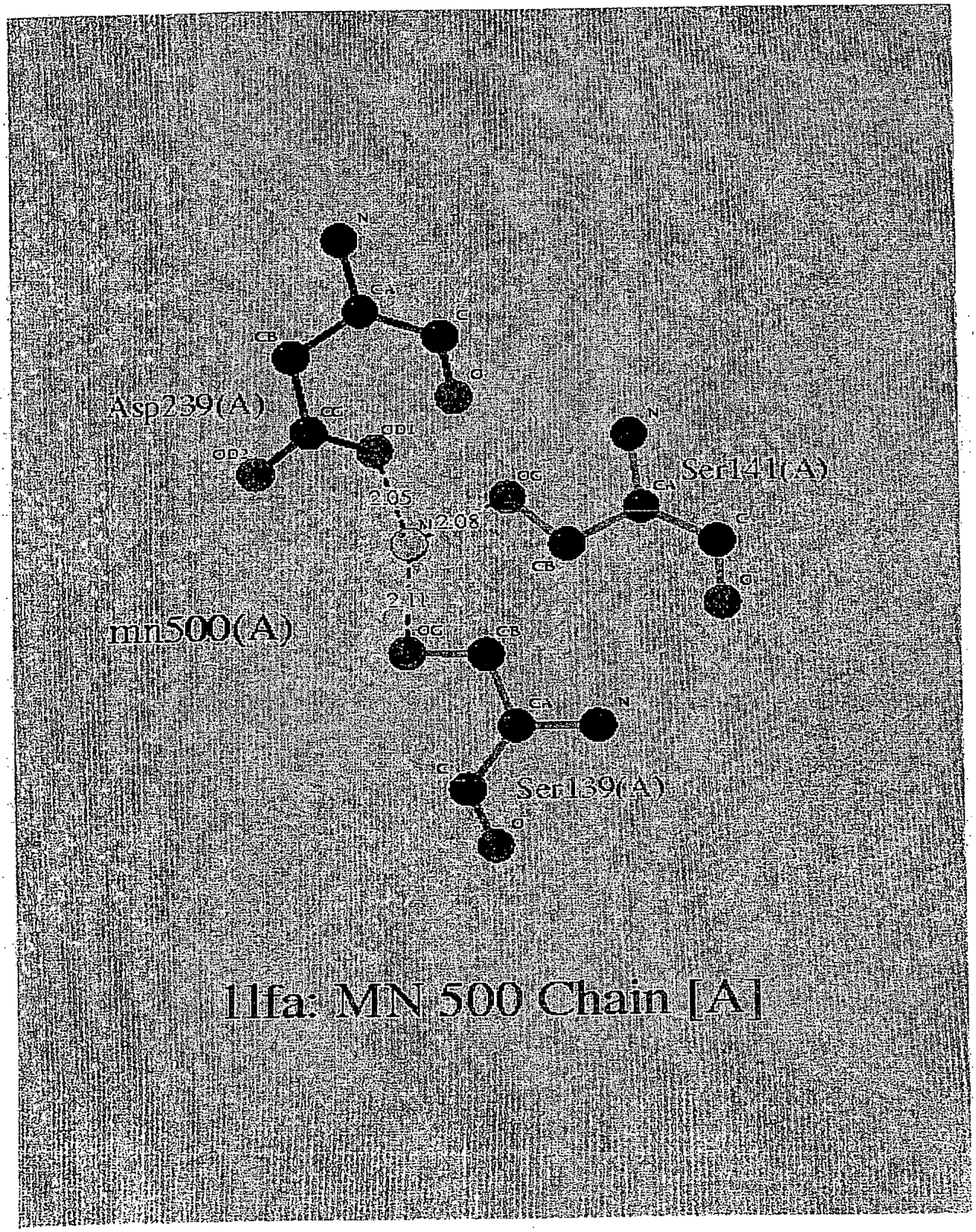

10 20 30 40 50 60 70
BAA15585.1 MTWFIDRRLNGKNKSMVNRQRFLRRYKAQIKQSISEAINKRSVTDVDSGESVSIPTEDISEPMFHQGRGGLRHRVHPGN
1LFA:A
80 90 100 110 120 130 140 150
BAA15585.1 DHFVQNDRIERPQGGGGGSGSGQGQASQDGEGQDEFVFQISKDEYLDLLFEDLALPNLKQNQQRQLTEYKTHRAGYTAN
1LFA:A
160 170 180 190 200 210 220 230
BAA15585.1 GVPANISVVRSLQNSLARRTAMTAGKRRELHALEENLAIISNSEPAQLLEEERLRKEIAELRAKIERVPFIDTFDLRYK
1LFA:A
240 250 260 270 280 290 300 310
BAA15585.1 NYEKRPDPSSQAVMFCLMDVSGSMDQSTKDMAKRFYILLYLFLSRTYKNVEVV----------YIRHHTQAKEVDEHEF
1LFA:A --------GNVDLVF-LFDGSMSLQPDEFQKILDFMKDVMKKLSNTSYQFAAVQFSTSYKTEFDFSDYVKRKDPDALLK
320 330 340 350 360 370 380 390
BAA15585.1 FYSQETGGTIVSSALKLMDEVVKERYNPAQWNIYAAQASDGDNWADDSPLCHEILAKKLLPVVRYYSYIEITRRA-HQT
1LFA:A HVKHMLLLTNTFGAINYVATEVFREELGARPDATKVLIIITDGEATD-----SGNIDAAKDIIRY--IIGIGKHFQTKE
400 410 420 430
BAA15585.1 LWREYEHLQSTFDNFAMQHIRDQDDIYPVFRELFHKQNATAKG
1LFA:A SQETLHKFASKPASEFVKILDTFEKLKDLFTE-LQKKIYVIE

Figure 8a



11fa: MN 500 Chain [A]

Figure 8B

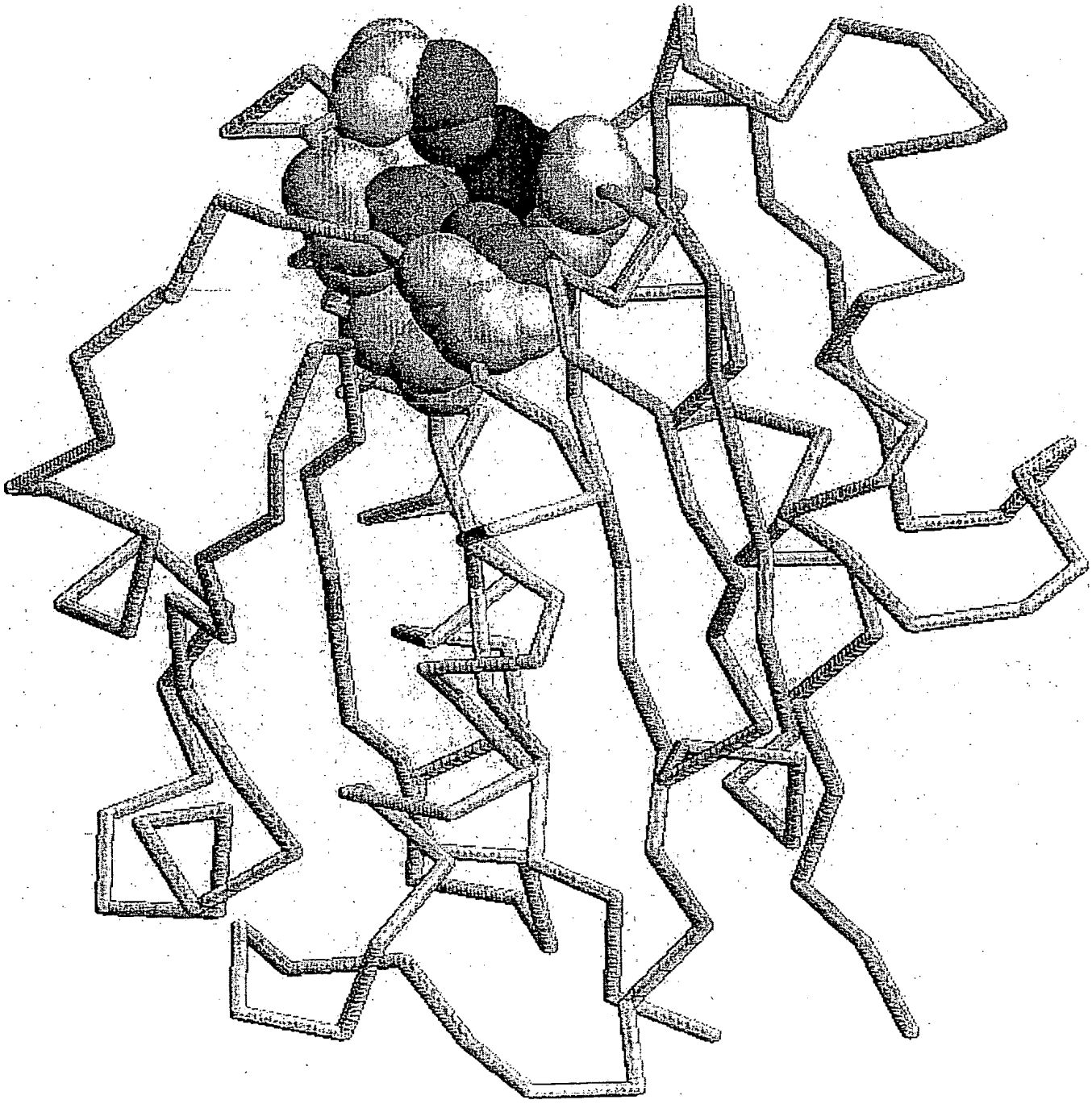
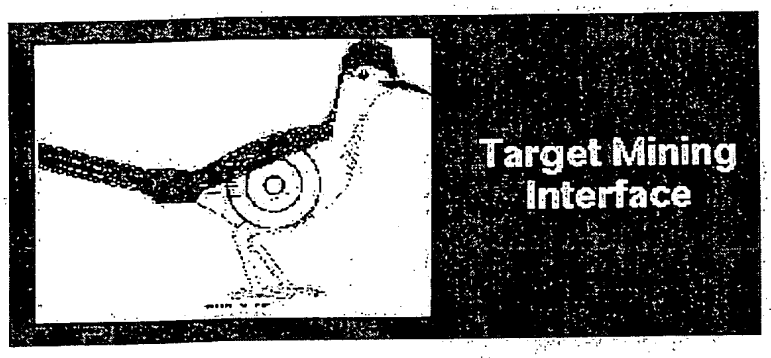


Figure 9

File Edit View Go Communicator Help

Bookmarks Location: http://victoria.inpharmatica.co.uk/~volker/BPD3target.html



Target Mining Interface

inpharmatica

Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): 1AOX and chain (e.g. B): A

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release: DEVT9 - BPD3

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

None

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

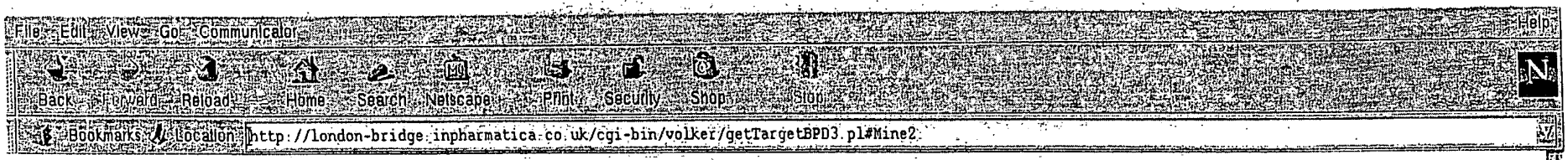


Figure 10A

2) 82 additional hits identified by both, Genome Threader and PSI-BLAST:

Combined Genome Threader and PSI-Blast output: PSI-BLAST values are shown in maroon!

Add2list	BPD links	WWW link	Title	Organism	Div.	%ID (GT,PSI)	Query rgn. (GT,PSI)	Target rgn. (GT,PSI)	Aln. score (GT)	Conf. (GT)	1st Iter. (PSI)	Best Iter. (PSI)	Best E-value (PSI)
☐	P56199 drill through Top50BlastHits Red.Seq.View	P56199	INTEGRIN ALPHA-1 (LAMININ AND COLLAGEN RECEPTOR) (VLA-1) (CD49A).	Homo sapiens (Human).	PRI	51.8%, 51% unmaskedSW	2-200, 2-200	139-337, 139-337	487	100% unmaskedGT	1	2	2E-73
☐	P17301 drill through Top50BlastHits Red.Seq.View	P17301	PLATELET MEMBRANE GLYCOPROTEIN IA PRECURSOR (GPIA) (COLLAGEN RECEPTOR) (INTEGRIN ALPHA-2) (VLA-2 ALPHA CHAIN) (CD49B).	Homo sapiens (Human).	PRI	99.5%, 100% unmaskedSW	2-200, 2-200	169-367, 169-367	471	100% unmaskedGT	1	1	1E-111
☐	AAF01258.1 drill through Top50BlastHits Red.Seq.View	AAF01258.1	integrin alpha-11 subunit precursor	Homo sapiens	PRI	45.2%, 45% unmaskedSW	2-200, 2-200	159-355, 159-355	469	100% unmaskedGT	1	2	4E-68
☐	AAD51919.2 drill through Top50BlastHits Red.Seq.View	AAD51919.2	integrin alpha 11 subunit precursor	Homo sapiens	PRI	45.2%, 45% unmaskedSW	2-200, 2-200	159-355, 159-355	469	100% unmaskedGT	1	2	4E-68
☐	Q99715 drill through Top50BlastHits Red.Seq.View	Q99715	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR.	Homo sapiens (Human).	PRI	28.1%, 31% unmaskedSW	2-196, 7-189	133-318, 2323-2495	437	100% unmaskedGT	1	2	2E-42
☐	AAA59491.1 drill through Top50BlastHits Red.Seq.View	AAA59491.1	Not given	Homo sapiens	PRI	27.1%, 27% unmaskedSW	2-200, 2-200	144-338, 144-338	435	100% unmaskedGT	1	2	1E-60
☐	AAA59544.1 drill through Top50BlastHits Red.Seq.View	AAA59544.1	Not given	Homo sapiens	PRI	27.1%, 27% unmaskedSW	2-200, 2-200	144-338, 144-338	435	100% unmaskedGT	1	2	1E-60
☐	CAB71222.1 drill through Top50BlastHits Red.Seq.View	CAB71222.1	dJ238D15.1 (collagen, type XII, alpha 1)	Homo sapiens	PRI	26.9%, 26% unmaskedSW	5-195, 156-201	293-472, 1581-1626	424	100% unmaskedGT	2	3	5E-6
☐	P21941 drill through Top50BlastHits Red.Seq.View	P21941	CARTILAGE MATRIX PROTEIN PRECURSOR (MATRILIN-1).	Homo sapiens (Human).	PRI	26.8%, 26% unmaskedSW	3-198, 3-198	271-455, 37-224	424	100% unmaskedGT	1	2	3E-46
☐	AAC01506.1 drill through Top50BlastHits Red.Seq.View	AAC01506.1	type XII collagen	Homo sapiens	PRI	27.6%, 30% unmaskedSW	2-195, 2-195	133-318, 133-318	422	100% unmaskedGT	1	2	3E-50
☐	AAC31952.1 drill through Top50BlastHits Red.Seq.View	AAC31952.1	integrin subunit alpha 10 precursor	Homo sapiens	PRI	46.5%, 46% unmaskedSW	2-199, 2-199	162-359, 162-359	419	100% unmaskedGT	1	2	5E-63
☐	CAA07569.1 drill through Top50BlastHits	CAA07569.1	matrilin-4	Homo sapiens	PRI	26%, 28% unmaskedSW	1-199, 2-200	335-529, 28-223	411	100% unmaskedGT	1	2	1E-47

										GT confidence			
☐	Top50BlastHits Red.Seq.View	BAA91707.1	Not given	sapiens	PRI	unmaskedSW	80-195	36-141	142	unmaskedGT	3	5	2E-8
☐	AAA36795.1 drill through Top50BlastHits Red.Seq.View	AAA36795.1	undulin 2	Homo sapiens	PRI	34.7%, 21% unmaskedSW	2-46, 139-196	3-51, 37-93	121	100% unmaskedGT	2	2	8E-6
☐	AAD40367.1 drill through Top50BlastHits Red.Seq.View	AAD40367.1	calcium-activated chloride channel-2	Homo sapiens	PRI	15%, 15% unmaskedSW	8-117, 8-117	312-421, 312-421	117	100% unmaskedGT	5	5	1E-4
☐	CAA67559.1 drill through Top50BlastHits Red.Seq.View	CAA67559.1	collagen VI-alpha-1 chain	Homo sapiens	PRI	21.1%, 21% unmaskedSW	108-197, 108-197	7-92, 7-92	114	100% unmaskedGT	3	3	5E-7
→	AAC76768.1 drill through Top50BlastHits Red.Seq.View	AAC76768.1	orf, hypothetical protein	Escherichia coli	BCT	13.2%, 11% unmaskedSW	9-134, 7-186	267-384, 265-424	107	100% unmaskedGT	-4	-5	3E-24
☐	CAB43000.1 drill through Top50BlastHits Red.Seq.View	CAB43000.1	tellurite resistance	Escherichia coli	BCT	12.9%, 14% unmaskedSW	8-192, 2-181	214-398, 208-380	102	100% unmaskedGT	-6	-7	6E-22
☐	AAA60114.1 drill through Top50BlastHits Red.Seq.View	AAA60114.1	platelet membrane glycoprotein IIb	Homo sapiens	PRI	14%, 13% unmaskedSW	157-199, 157-199	145-187, 145-187	111	98.52% unmaskedGT reverse Hit	-7	-8	8E-5

608 out of these 632 PSI-BLAST matches were identified using 'positive iterations':

P17301 ,1AOX:B ,AAB59255.1 ,CAA82877.1 ,CAA78125.1 ,P56199 ,CAA36384.1 ,BAA23160.1 ,1CK4:B ,1CK4:A ,1QC5:B ,1QC5:A ,AAF01258.1 ,AAD51913.2 ,AAC31352.1 ,CAA30479.1 ,AAA59544.1 ,AAB24821.1 ,1BHO:1 ,1BHO:2 ,1IDN:1 ,1BHQ:1 ,1BHQ:2 ,1IDN:2 ,AAB16869.1 ,1ILM ,CAA50063.1 ,P32018 ,AAA99719.1 ,BAA00701.1 ,Q99715 ,1IDO ,AAB38702.1 ,CAA06890.1 ,CAA06889.1 ,AAF21241.1 ,CAA07569.1 ,AAB38547.1 ,CAB46380.1 ,P51942 ,P20701 ,P05099 ,AAC31672.1 ,1DGO:A ,Q91145 ,CAB71222.1 ,1LFA:A ,1EDO:B ,1EDO:A ,1EDP:B ,1EDP:A ,Q99715 ,BAA00701.1 ,1LFA:B ,AAA99719.1 ,AAD25884.1 ,AAA39426.1 ,AAD25885.1 ,Q99715 ,1EDN ,AAC23663.1 ,CAB70853.1 ,AAC01506.1 ,CAA50063.1 ,P32018 ,AAA99719.1 ,AAC53163.1 ,CAA72402.1 ,AAF43107.1 ,P38570 ,AAD04919.1 ,AAB93766.1 ,BAA36034.1 ,AAC52142.1 ,BAA00701.1 ,CAA27972.1 ,AAB59512.1 ,AAA82929.1 ,AAC53163.1 ,CAB37653.1 ,CAA07569.1 ,CAA12110.1 ,CAB86185.1 ,CAA68283.1 ,CAB46380.1 ,CAB37847.1 ,AAB65641.1 ,AAC06229.1 ,AAF25353.1 ,AAB51548.1 ,AAD50539.1 ,CAB37672.1 ,CAB37665.1 ,CAA03885.1 ,CAB37852.1 ,P05099 ,AAB38702.1 ,AAB51556.1 ,AAB51557.1 ,P51942 ,CAB86200.1 ,CAB86203.1 ,CAB40275.1 ,CAB37666.1 ,CAB37856.1 ,AAB51554.1 ,CAB37651.1 ,AAB51553.1 ,AAB38798.1 ,AAA51620.1 ,AAB51550.1 ,CAA06889.1 ,AAC72024.1 ,CAB86202.1 ,CAA36267.1 ,AAF25351.1 ,AAD50540.1 ,CAB86227.1 ,AAD50544.1 ,AAD01701.1 ,AAD26126.1 ,AAB51546.1 ,AAD26124.1 ,AAB21220.2 ,AAF42927.1 ,CAB86184.1 ,AAB65642.1 ,CAB86188.1 ,AAC32844.1 ,CAB37670.1 ,CAB37668.1 ,AAD50542.1 ,AAB38800.1 ,AAD26125.1 ,AAB03226.1 ,AAC32845.1 ,AAF25354.1 ,CAB65984.1 ,CAB65984.1 ,CAB37843.1 ,AAB38802.1 ,AAB51552.1 ,CAB37667.1 ,AAB51549.1 ,AAC32846.1 ,Q99715 ,AAB51555.1 ,AAD29428.1 ,AAB38799.1 ,AAF25352.1 ,CAB37857.1 ,AAB51551.1 ,AAA36794.1 ,AAA28092.2 ,CAB37849.1 ,AAB65643.1 ,AAB38797.1 ,AAF23492.1 ,AAD01978.1 ,AAA03201.1 ,AAB51558.1 ,CAB37846.1 ,AAB51545.1 ,AAF42926.1 ,AAD50543.1 ,AAA03201.1 ,AAA03201.1 ,AAA99719.1 ,CAA36267.1 ,AAF42925.1 ,AAA03201.1 ,CAA71532.1 ,AAA96311.1 ,AAD01978.1 ,AAC23667.1 ,AAC62253.1 ,CAA27972.1 ,AAB59512.1 ,1ATZ:B ,1AO3:B ,1AO3:A ,CAA36267.1 ,AAA03201.1 ,AAD26127.1 ,AAD01978.1 ,CAA36267.1 ,AAD01701.1 ,AAD04319.1 ,1ATZ:A ,AAB93766.1 ,AAD01978.1 ,AAD40483.1 ,CAA36267.1 ,CAA97806.1 ,CAA36267.1 ,AAF01565.1 ,CAA36267.1 ,AAD01978.1 ,AAC23667.1 ,AAB63303.1 ,AAC62253.1 ,Q43405 ,AAA29076.1 ,AAC39949.1 ,AAA03201.1 ,AAF19243.1 ,AAC23668.1 ,AAC06229.1 ,AAD01978.1 ,BAA23627.1 ,AAA03201.1 ,AAA03201.1 ,AAC23667.1 ,CAA36267.1 ,AAD03350.1 ,AAC23667.1 ,AAC73682.1 ,AAD01978.1 ,CAA36267.1 ,AAC68807.1 ,BAA82322.1 ,CAA47744.1 ,CAA98248.1 ,BAA00701.1 ,AAC47957.1 ,AAA50715.1 ,AAC19251.1 ,AAB52325.1 ,AAF19243.1 ,BAA82322.1 ,AAC02556.1 ,AAB66593.1 ,AAD01978.1 ,AAC23667.1 ,AAA58965.1 ,Q02388 ,CAB37672.1 ,AAD01978.1 ,AAC23667.1 ,CAB37849.1 ,AAB51551.1 ,AAF25353.1 ,AAB38799.1 ,CAB37670.1 ,AAB38797.1 ,CAB86200.1 ,CAB58678.1 ,1AUQ ,1OAK:A ,AAB51558.1 ,CAB37653.1 ,AAB51557.1 ,AAB59512.1 ,CAA27972.1 ,AAD50539.1 ,CAB86202.1 ,AAD50540.1 ,AAD50541.1 ,CAB37667.1 ,AAB65641.1 ,CAB37665.1 ,AAB65642.1 ,AAB51548.1 ,CAB86188.1 ,AAB51547.1 ,CAB37843.1 ,AAF42927.1 ,AAF25354.1 ,AAF42925.1 ,AAD26125.1 ,CAB86184.1 ,AAB51555.1 ,AAD26126.1 ,AAD26127.1 ,AAB51550.1 ,CAB86185.1 ,CAB37857.1 ,AAB51552.1 ,CAB86203.1 ,AAC06229.1 ,CAB37668.1 ,BAA31172.1 ,CAB37846.1 ,AAA29773.1 ,CAA31440.1 ,AAD04319.1 ,CAB37852.1 ,AAA96311.1 ,CAB86227.1 ,AAB93766.1 ,AAB51559.1 ,AAA29774.1 ,AAA29776.1 ,AAA29770.1 ,AAA29772.1 ,CAB37847.1 ,AAD50543.1 ,AAB51560.1 ,AAB38798.1 ,AAC97484.1 ,BAA31171.1 ,BAA31173.1 ,BAA31174.1 ,BAA31182.1 ,BAA31190.1 ,BAA31188.1 ,BAA31178.1 ,AAA29778.1 ,BAA31167.1 ,BAA31193.1 ,BAA31177.1 ,BAA31176.1 ,BAA31195.1 ,BAA31194.1 ,BAA31169.1 ,BAA31175.1 ,AAC18657.1 ,BAA31183.1 ,AAA29771.1 ,BAA31187.1 ,AAB38802.1 ,AAC47463.1 ,AAA29767.1 ,AAA29775.1 ,BAA31170.1 ,AAB21220.2 ,AAB51549.1 ,AAA03201.1 ,CAB37856.1 ,Q43405 ,AAC39949.1 ,AAB51554.1 ,AAC46961.1 ,AAB38801.1 ,CAB38978.1 ,AAC47462.1 ,CAA33891.1 ,AAA35618.1 ,AAA35619.1 ,AAD50544.1 ,CAA46541.1 ,CAA73140.1 ,AAC32844.1 ,AAF25352.1 ,AAB34889.1 ,AAC23667.1 ,AAF25351.1 ,AAC32845.1 ,CAB37651.1 ,AAD26124.1 ,AAB63302.1 ,AAB51546.1 ,AAC84162.1 ,CAB37666.1 ,AAB51556.1 ,CAA73152.1 ,CAA33144.1 ,CAA39881.1 ,AAA82929.1 ,AAA37380.1 ,AAA37381.1 ,AAB51545.1 ,AAB97607.1 ,BAA34707.1 ,Q04857 ,BAA07526.1 ,AAB51547.1 ,AAB51553.1 ,AAF69610.1 ,AAD50542.1 ,CAA33889.1 ,CAA67576.1 ,BAA02763.1 ,CAB03057.1 ,AAF42926.1 ,CAA51389.1 ,AAC32846.1 ,BAA06173.1 ,AAC47957.1 ,CAA33888.1 ,Q04857 ,BAA82322.1 ,AAB59954.1 ,AAC47461.1 ,CAB03055.1 ,CAB03054.1 ,AAB65643.1 ,BAA88322.1 ,CAB03056.1 ,CAB03059.1 ,CAB38978.1 ,AAC46961.1 ,BAA82322.1 ,CAB71222.1 ,AAC84160.1 ,AAB59954.1 ,CAB03058.1 ,CAA73153.1 ,AAB24637.1 ,AAB38800.1 ,AAB59954.1 ,CAB46421.1 ,CAA46541.1 ,BAA02853.1 ,CAA33891.1 ,CAA67576.1 ,CAA73153.1 ,Q04857 ,CAA88985.1 ,CAA73152.1 ,AAC46572.1 ,AAF00021.1 ,CAA33144.1 ,CAA39881.1 ,AAC24429.1 ,CAB75310.1 ,AAC23667.1 ,CAB38978.1 ,AAC46961.1 ,CAB02952.1 ,AAB51560.1 ,CAA33890.1 ,AAA58965.1 ,Q02388 ,CAA36267.1 ,AAA03201.1 ,AAF70042.1 ,CAB03680.1 ,CAA33889.1 ,AAD20859.1 ,CAA33144.1 ,CAA73156.1 ,AAA03201.1 ,CAB38978.1 ,AAC65240.1 ,BAA23627.1 ,CAA46541.1 ,P34333 ,AAD50541.1 ,CAA91487.1 ,AAD01978.1 ,BAA82322.1 ,AAC17639.1 ,AAC23667.1 ,CAA73156.1 ,AAD16289.1 ,AAB04590.2 ,CAA61109.1 ,AAB38801.1 ,CAB38978.1 ,AAC65016.1 ,AAB63363.1 ,AAA36795.1 ,AAC46961.1 ,AAC46961.1 ,AAC46572.1 ,AAF12450.1 ,AAD50541.1 ,AAA81430.1 ,AAA50715.1 ,CAB38978.1 ,AAC46961.1 ,AAC13251.1 ,AAA36795.1 ,CAA36267.1 ,Q91145 ,Q61703 ,BAA13939.1 ,P19823 ,CAA72308.1 ,CAA30160.1 ,Q61704 ,CAA72155.1 ,BAA78416.1 ,BAA13940.1 ,AAC79982.1 ,Q63416 ,AAC35003.1 ,AAC36289.1 ,P54290 ,AAB50142.1 ,AAB50141.1 ,AAB50140.1 ,AAB50139.1 ,AAB50138.1 ,CAA49279.1 ,CAA45188.1 ,AAA81562.1 ,AAF03259.1 ,BAA34706.1 ,CAA72309.1 ,BAA13938.1 ,AAC48511.1 ,AAB46821.1 ,P54289 ,AAF12731.1 ,BAA07632.1 ,CAA49841.1 ,AAB86529.1 ,AAB86531.1 ,AAB86530.1 ,CAB86192.1 ,AAC70914.1 ,CAB86193.1 ,CAA88984.1 ,AAD48398.1 ,Q55874 ,AAC25786.1 ,CAB77598.1 ,P76481 ,AAF53476.1 ,AAF44957.1 ,BAA12207.1 ,AAF13350.1 ,CAB10840.1 ,CAA88986.1 ,AAA83291.1 ,AAC79610.1 ,CAA88991.1 ,AAB63940.1 ,BAA08371.1 ,CAA88988.1 ,CAA91030.1 ,AAA36225.2 ,CAB08456.1 ,AAD30858.1 ,AAB19093.1 ,AAC72024.1 ,CAA16338.1 ,CAA94611.1 ,AAC05096.1 ,CAB59466.1 ,BAA80016.1 ,AAC79447.1 ,AAF39956.1 ,AAF10487.1 ,AAC65716.1 ,CAB11610.1 ,CAA10812.1 ,BAA91707.1 ,AAD34763.1 ,CAB03681.1 ,CAA67559.1 ,AAD03499.2 ,CAB38978.1 ,AAC46961.1 ,AAB51559.1 ,AAC95429.1 ,CAA09423.1 ,BAA33743.1 ,CAB75878.1 ,CAB75962.1 ,AAF58335.1 ,CAA37671.1 ,AAF53505.1 ,AAF44988.1 ,CAB71206.1 ,AAF27140.1 ,CAA47439.1 ,AAC83699.1 ,CAB76889.1 ,BAA25490.1 ,BAA20346.1 ,AAF22895.1 ,CAB69268.1 ,AAD52031.1 ,BAA16787.1 ,CAB58179.1 ,O22437 ,AAC83699.1 ,BAA18152.1 ,AAF69021.1 ,AAC75134.1 ,AAA60558.1 ,AAB85061.1 ,O50313 ,AAC75182.1 ,CAA16691.1 ,CAB80454.1 ,AAD41583.1 ,AAF58612.1 ,AAA27969.1 ,BAA76532.1 ,AAD31495.2 ,AAC44737.1 ,CAA77537.1 ,O34845 ,AAC84032.1 ,AAC62491.1 ,AAD15576.1 ,AAD47250.1 ,CAB13830.1 ,CAA90091.1 ,AAB98810.1 ,BAA79575.1 ,Q60284 ,BAA61542.1 ,AAF11936.1 ,AAD40367.1 ,AAB89815.1 ,AAA68973.1 ,BAA10368.1 ,AAF23117.1 ,AAF27914.1 ,CAB52078.1 ,BAA20761.1 ,AAD42198.2 ,AAB84957.1 ,

24 out of these 632 PSI-BLAST matches were identified using 'negative iterations':

BAA81233.1 ,CAA12066.1 ,CAB50193.1 ,CAB43000.1 ,CAB01991.1 ,CAA17374.1 ,BAA07331.1 ,AAC76768.1 ,AAB86281.1 ,CAB76247.1 ,AAC45374.1 ,AAB80774.1 ,AAA86852.2 ,AAB38912.1 ,AAA33386.1 ,AAA33386.1 ,AAC16931.1 ,AAF57485.1 ,BAA86551.1 ,AAB41649.1 ,AAA60114.1 ,AAC51323.1 ,AAC52946.1 ,Q01443 ,

Figure 11 electronic

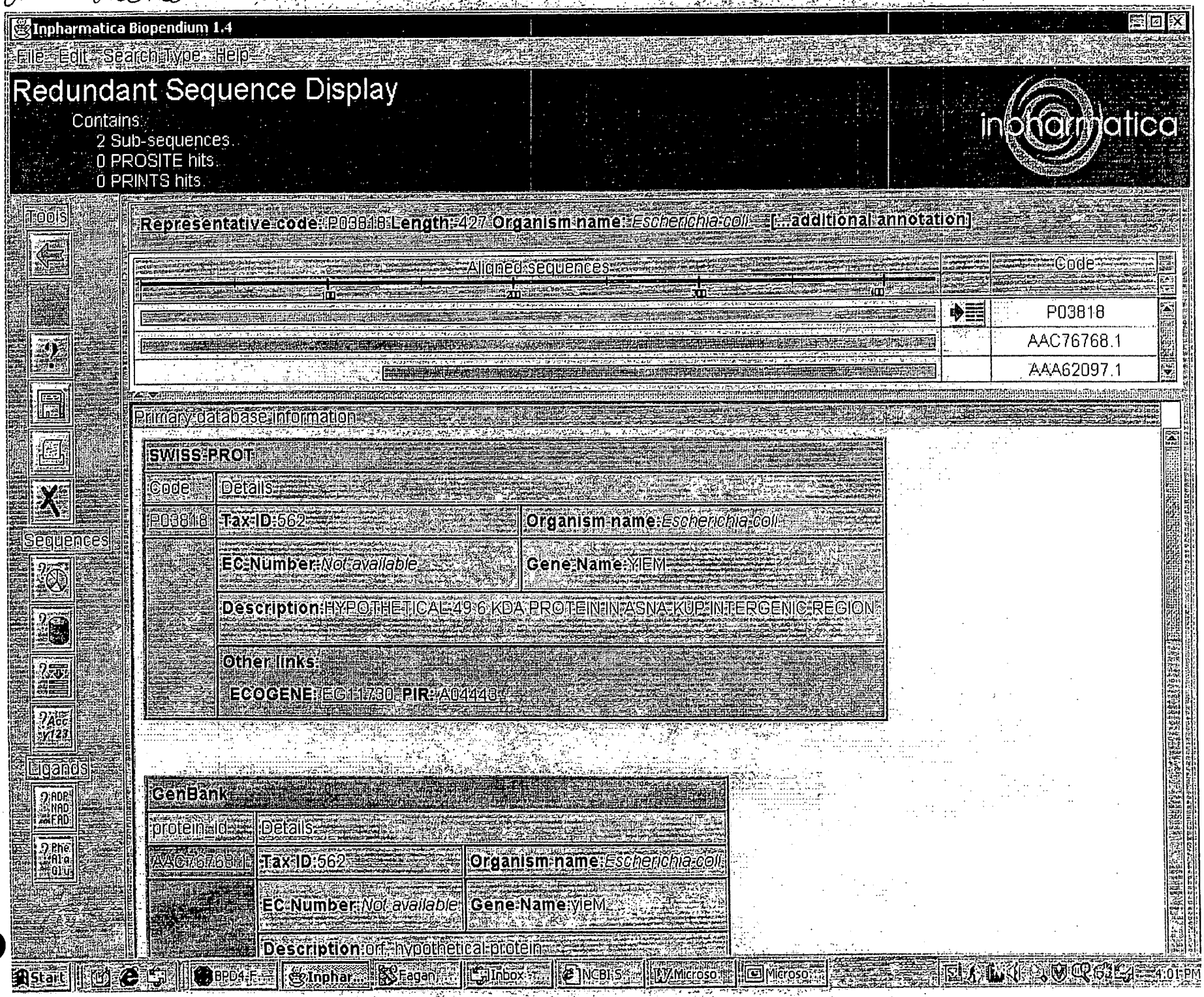
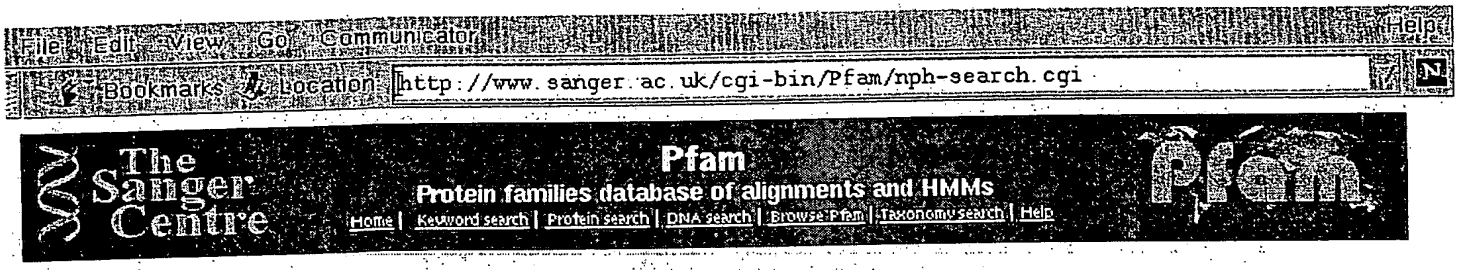


Figure 12



Results for gi|2367274|gb|AAC76768.1|

There were no matches to Pfam-A (including borderline matches) for gi|2367274|gb|AAC76768.1|

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_15204	204	408	2.4e-108	Align

[427 residues]

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families: Hypertext linked to swisspfam

Query gi|2367274|gb|AAC76768.1|/204-408 matching Pfam-B_15204

```
                YIEM_ECOLI  204 DILRLLPPELATLGITELEYEFYRRLVEKQLLTYRLHGESWREKVIERPV  253
                            DILRLLPPELATLGITELEYEFYRRLVEKQLLTYRLHGESWREKVIERPV
gi|2367274|gb|AAC76768.1|  204 DILRLLPPELATLGITELEYEFYRRLVEKQLLTYRLHGESWREKVIERPV  253

                YIEM_ECOLI  254 VHKDYDEQPRGPFIVCVDTSGSMGGFNEQCAKAFCLALMRIALAENRRCY  303
                            VHKDYDEQPRGPFIVCVDTSGSMGGFNEQCAKAFCLALMRIALAENRRCY
gi|2367274|gb|AAC76768.1|  254 VHKDYDEQPRGPFIVCVDTSGSMGGFNEQCAKAFCLALMRIALAENRRCY  303

                YIEM_ECOLI  304 IMLFSTEIVRYELSGPQGIEQAIRFLSQQFRGGTDLASCFRAIMERLQSR  353
                            IMLFSTEIVRYELSGPQGIEQAIRFLSQQFRGGTDLASCFRAIMERLQSR
gi|2367274|gb|AAC76768.1|  304 IMLFSTEIVRYELSGPQGIEQAIRFLSQQFRGGTDLASCFRAIMERLQSR  353

                YIEM_ECOLI  354 EWFDADAVVISDFIAQRLPDDVTSKVKELQRVHQMRFHAVAMSAMGKPGI  403
                            EWFDADAVVISDFIAQRLPDDVTSKVKELQRVHQMRFHAVAMSAMGKPGI
gi|2367274|gb|AAC76768.1|  354 EWFDADAVVISDFIAQRLPDDVTSKVKELQRVHQMRFHAVAMSAMGKPGI  403

                YIEM_ECOLI  404 MRIFD 408
                            MRIFD
gi|2367274|gb|AAC76768.1|  404 MRIFD 408
```

Align to family

If you think there is anything wrong with this script, please contact Pfam

Figure 13

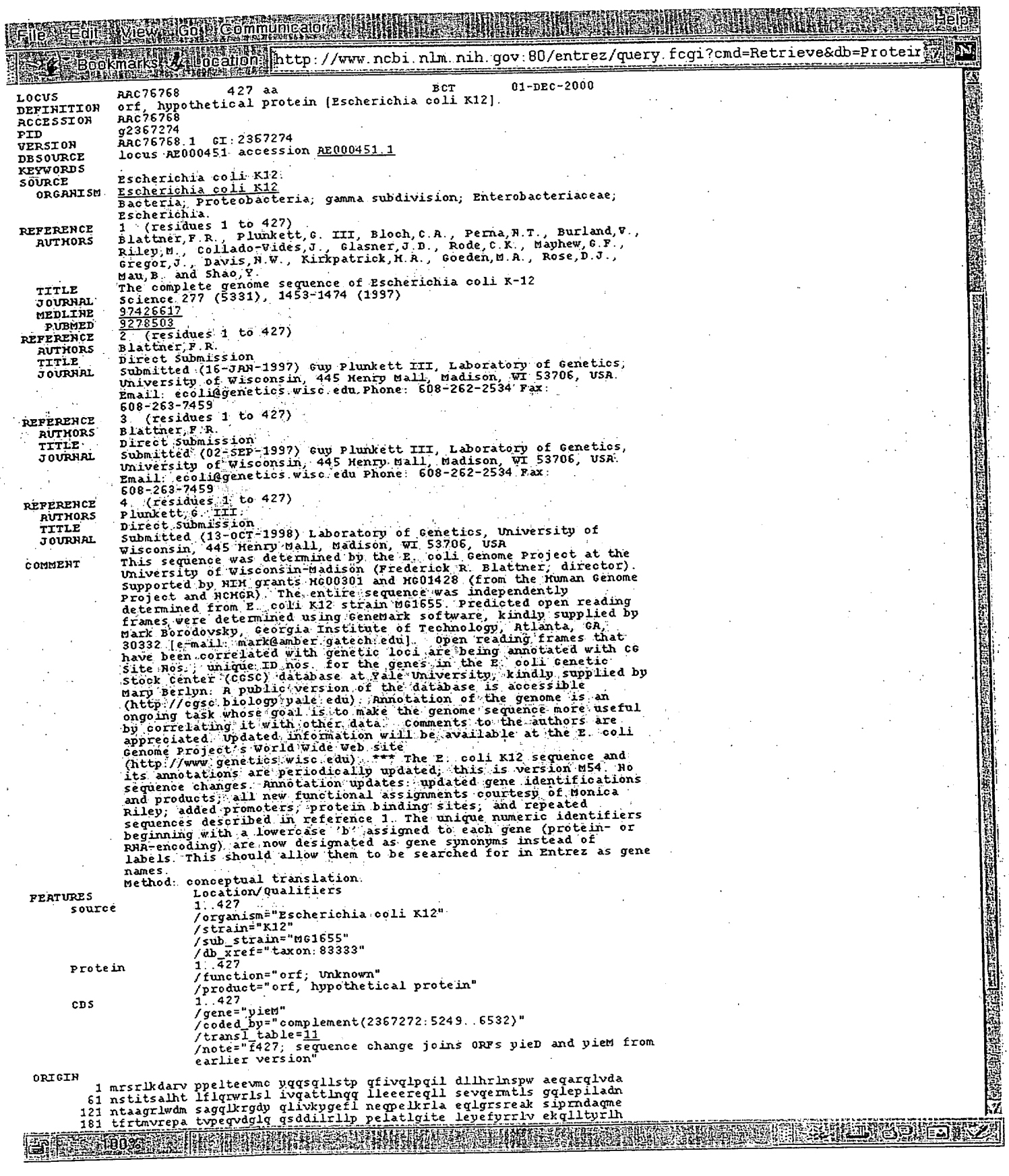
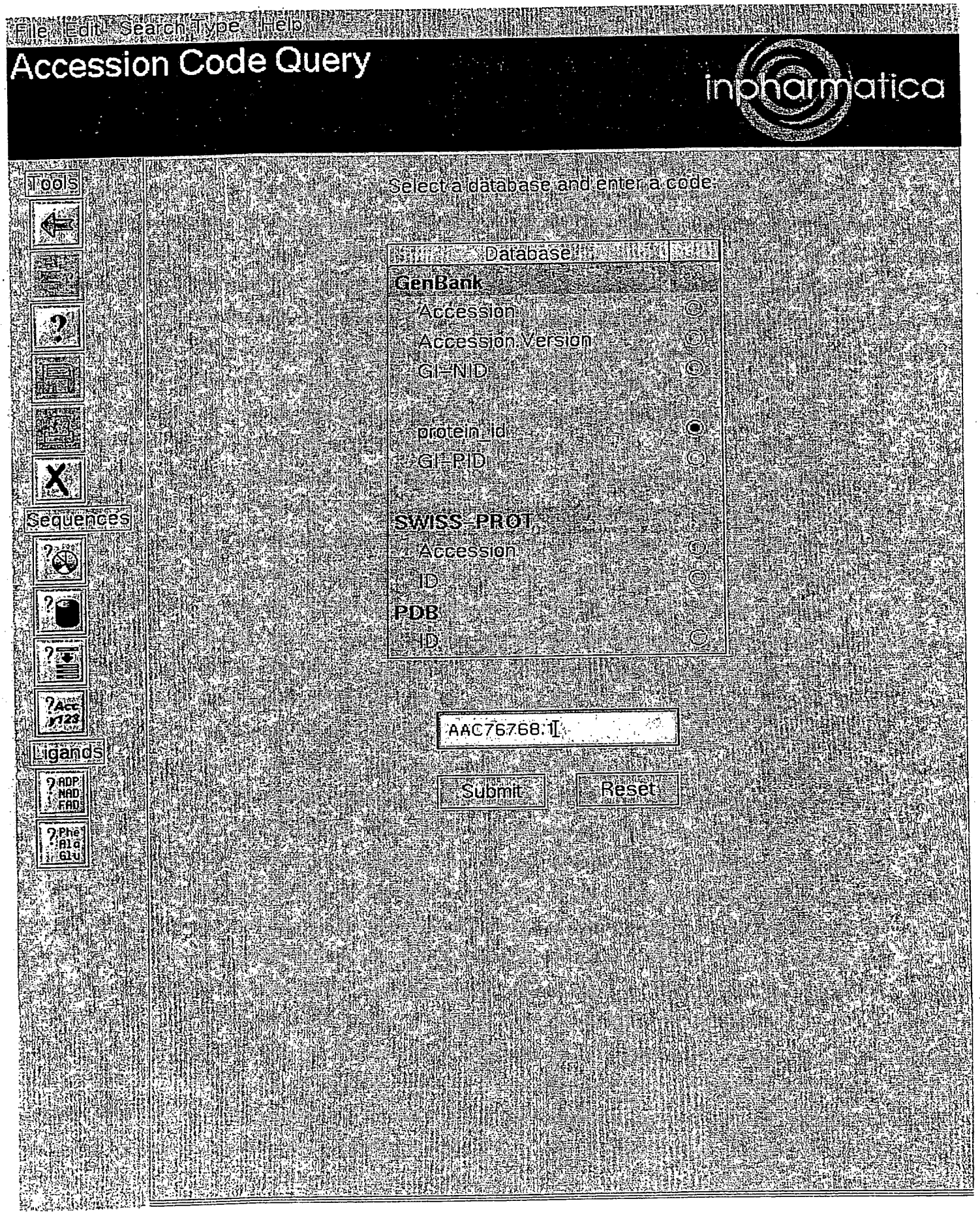

File Edit View Go Communicator Help
Bookmarks Location: http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=Protein
LOCUS AAC76768 427 aa BCT 01-DEC-2000
DEFINITION orf, hypothetical protein [Escherichia coli K12].
ACCESSION AAC76768
PID g2367274
VERSION AAC76768.1 GI:2367274
DBSOURCE locus AE000451 accession AE000451.1
KEYWORDS
SOURCE Escherichia coli K12.
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (residues 1 to 427)
AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (residues 1 to 427)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 3 (residues 1 to 427)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 4 (residues 1 to 427)
AUTHORS Plunkett,G. III
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and HCMGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos.; unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.
Method: conceptual translation.
FEATURES Location/Qualifiers
source 1..427
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
Protein 1..427
/function="orf; Unknown"
/product="orf, hypothetical protein"
CDS 1..427
/gene="yieM"
/coded_by="complement(2367272:5249..6532)"
/transl_table=11
/note="f427; sequence change joins ORFs yieD and yieM from earlier version"
ORIGIN
1 mrsrlkdarv ppelteevmc yqqsqllstp qfivqlpqil dllhrlnspw aeqarqlvda
61 nstitsalht lflqrwrlsl ivqattlnqq lleeereqll sevqermtls gqlepiladn
121 ntaagrlwdm sagqlkrgdy qlivkygefl neqpelkrla eqlgrsreak siprndaqme
181 tfrtmvrepa tvpeqvdglq qsddilrllp pelatlgite leyefyrrlv ekqlltyrlh
100%

Figure 14A


File Edit Search Type Help
Accession Code Query
inpharmatica
Tools
Sequences
Ligands
ADP NAD FAD
Phe Ala Glu
Select a database and enter a code:
Database
GenBank
Accession
Accession.Version
GI_NID
protein_id
GI_PID
SWISS_PROT
Accession
ID
PDB
ID
AAC76768.1
Submit
Reset

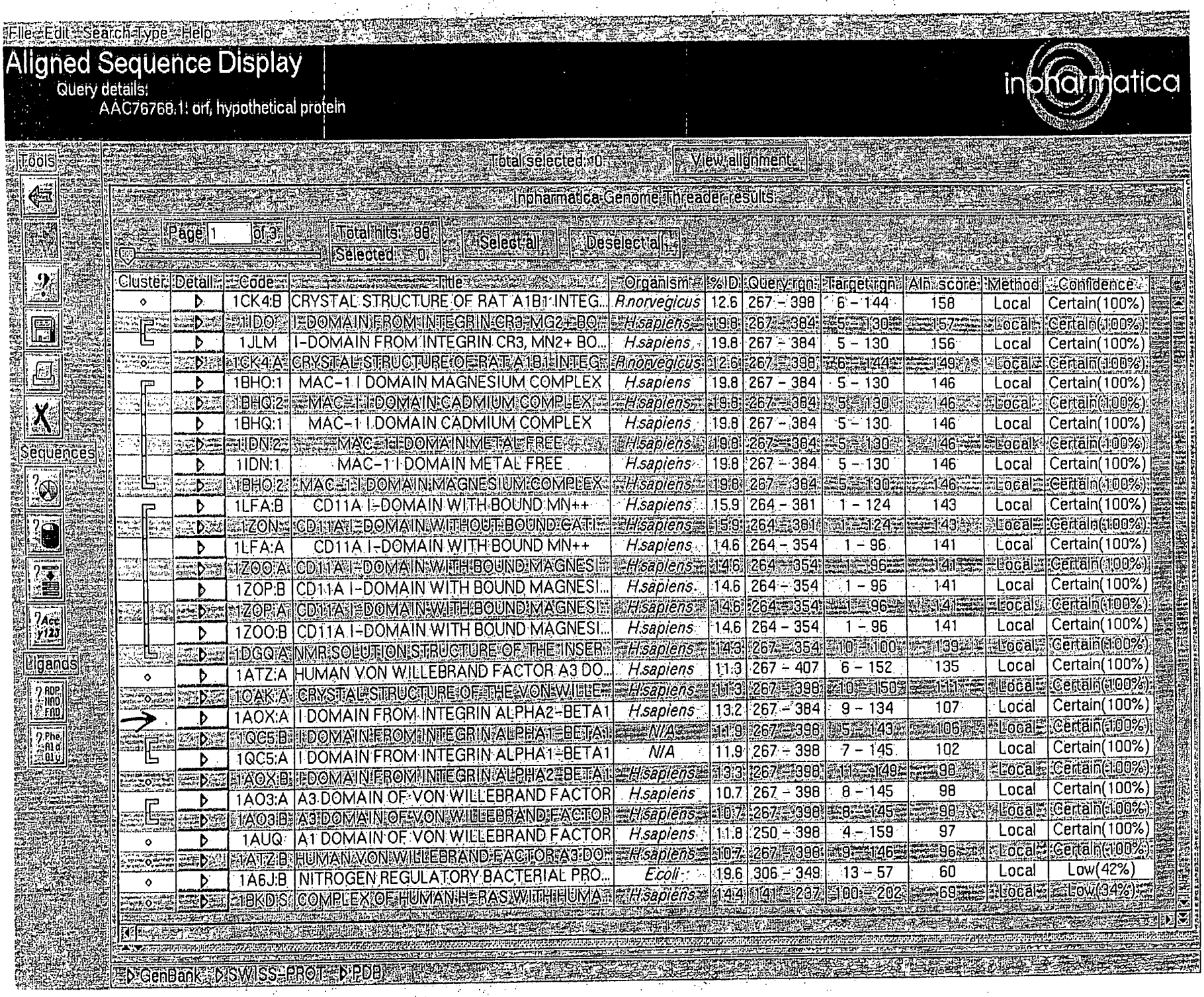

File Edit Search Type Help
Aligned Sequence Display
Query details:
AAC76768.1: orf, hypothetical protein
inpharmatica
Tools
Sequences
Ligands
Total selected: 0
View alignment
Inpharmatica Genome Threader results
Page 1 of 3
Total hits: 88
Selected: 0
Select all
Deselect all
Cluster | Detail | Code | Title | Organism | %ID | Query rgn | Target rgn | Aln score | Method | Confidence
1CK4:B | CRYSTAL STRUCTURE OF RAT A1B1 INTEG... | R.norvegicus | 12.6 | 267 – 398 | 6 – 144 | 158 | Local | Certain(100%)
1IDO | I-DOMAIN FROM INTEGRIN CR3, MG2+ BO... | H.sapiens | 19.8 | 267 – 384 | 5 – 130 | 157 | Local | Certain(100%)
1JLM | I-DOMAIN FROM INTEGRIN CR3, MN2+ BO... | H.sapiens | 19.8 | 267 – 384 | 5 – 130 | 156 | Local | Certain(100%)
1CK4:A | CRYSTAL STRUCTURE OF RAT A1B1 INTEG... | R.norvegicus | 12.6 | 267 – 398 | 6 – 144 | 149 | Local | Certain(100%)
1BHO:1 | MAC-1 I DOMAIN MAGNESIUM COMPLEX | H.sapiens | 19.8 | 267 – 384 | 5 – 130 | 146 | Local | Certain(100%)
1BHQ:2 | MAC-1 I-DOMAIN CADMIUM COMPLEX | H.sapiens | 19.8 | 267 – 384 | 5 – 130 | 146 | Local | Certain(100%)
1BHQ:1 | MAC-1 I DOMAIN CADMIUM COMPLEX | H.sapiens | 19.8 | 267 – 384 | 5 – 130 | 146 | Local | Certain(100%)
1IDN:2 | MAC-1 I-DOMAIN METAL FREE | H.sapiens | 19.8 | 267 – 384 | 5 – 130 | 146 | Local | Certain(100%)
1IDN:1 | MAC-1 I DOMAIN METAL FREE | H.sapiens | 19.8 | 267 – 384 | 5 – 130 | 146 | Local | Certain(100%)
1BHO:2 | MAC-1 I DOMAIN MAGNESIUM COMPLEX | H.sapiens | 19.8 | 267 – 384 | 5 – 130 | 146 | Local | Certain(100%)
1LFA:B | CD11A I-DOMAIN WITH BOUND MN++ | H.sapiens | 15.9 | 264 – 381 | 1 – 124 | 143 | Local | Certain(100%)
1ZON | CD11A I-DOMAIN WITHOUT BOUND CATI... | H.sapiens | 15.9 | 264 – 381 | 1 – 124 | 143 | Local | Certain(100%)
1LFA:A | CD11A I-DOMAIN WITH BOUND MN++ | H.sapiens | 14.6 | 264 – 354 | 1 – 96 | 141 | Local | Certain(100%)
1ZOO:A | CD11A I-DOMAIN WITH BOUND MAGNESI... | H.sapiens | 14.6 | 264 – 354 | 1 – 96 | 141 | Local | Certain(100%)
1ZOP:B | CD11A I-DOMAIN WITH BOUND MAGNESI... | H.sapiens | 14.6 | 264 – 354 | 1 – 96 | 141 | Local | Certain(100%)
1ZOP:A | CD11A I-DOMAIN WITH BOUND MAGNESI... | H.sapiens | 14.6 | 264 – 354 | 1 – 96 | 141 | Local | Certain(100%)
1ZOO:B | CD11A I-DOMAIN WITH BOUND MAGNESI... | H.sapiens | 14.6 | 264 – 354 | 1 – 96 | 141 | Local | Certain(100%)
1DGQ:A | NMR SOLUTION STRUCTURE OF THE INSER... | H.sapiens | 14.3 | 267 – 354 | 10 – 100 | 139 | Local | Certain(100%)
1ATZ:A | HUMAN VON WILLEBRAND FACTOR A3 DO... | H.sapiens | 11.3 | 267 – 407 | 6 – 152 | 135 | Local | Certain(100%)
1OAK:A | CRYSTAL STRUCTURE OF THE VON WILLE... | H.sapiens | 11.3 | 267 – 398 | 10 – 150 | 111 | Local | Certain(100%)
1AOX:A | I DOMAIN FROM INTEGRIN ALPHA2-BETA1 | H.sapiens | 13.2 | 267 – 384 | 9 – 134 | 107 | Local | Certain(100%)
1QC5:B | I DOMAIN FROM INTEGRIN ALPHA1-BETA1 | N/A | 11.9 | 267 – 398 | 5 – 143 | 106 | Local | Certain(100%)
1QC5:A | I DOMAIN FROM INTEGRIN ALPHA1-BETA1 | N/A | 11.9 | 267 – 398 | 7 – 145 | 102 | Local | Certain(100%)
1AOX:B | I DOMAIN FROM INTEGRIN ALPHA2-BETA1 | H.sapiens | 13.3 | 267 – 398 | 11 – 149 | 98 | Local | Certain(100%)
1AO3:A | A3 DOMAIN OF VON WILLEBRAND FACTOR | H.sapiens | 10.7 | 267 – 398 | 8 – 145 | 98 | Local | Certain(100%)
1AO3:B | A3 DOMAIN OF VON WILLEBRAND FACTOR | H.sapiens | 10.7 | 267 – 398 | 8 – 145 | 98 | Local | Certain(100%)
1AUQ | A1 DOMAIN OF VON WILLEBRAND FACTOR | H.sapiens | 11.8 | 250 – 398 | 4 – 159 | 97 | Local | Certain(100%)
1ATZ:B | HUMAN VON WILLEBRAND FACTOR A3 DO... | H.sapiens | 10.7 | 267 – 398 | 9 – 146 | 96 | Local | Certain(100%)
1A6J:B | NITROGEN REGULATORY BACTERIAL PRO... | E.coli | 19.6 | 306 – 349 | 13 – 57 | 60 | Local | Low(42%)
1BKD:S | COMPLEX OF HUMAN H-RAS WITH HUMA... | H.sapiens | 14.4 | 141 – 207 | 100 – 202 | 69 | Local | Low(34%)
GenBank SWISS-PROT PDB

Figure 143

Figure 14C

File Edit Search Type Help

Aligned Sequence Display

Query details:
AAC76768.1 orf, hypothetical protein

inpharmatica

Tools

Sequences

Ligands

Total selected: 0 | View alignment

Reverse Maximised PSI-BLAST results

Page 1 of 17 | Total hits: 506 | Selected: 0 | Select all | Deselect all

Cluster	Detail	Code	Title	Organism	%ID	Query rgn	Target rgn	Iter	E-value
◇	▷	BAA80016.1	452aa long hypothetical protein	*A.pernix*	28.0	156 – 407	202 – 431	2	2.00E-60
◇	▷	Q60384	HYPOTHETICAL PROTEIN MJ0077	*M.jannaschii*	22.0	146 – 408	122 – 361	3	2.00E-7
◇	▷	AAB98810.1	conserved hypothetical protein	*M.jannaschii*	23.0	147 – 410	170 – 420	3	2.00E-65
◇	▷	AAC79610.1	putative retroelement pol polyprotein	*A.thaliana*	19.0	168 – 404	144 – 403	3	2.00E-50
◇	▷	Q55874	HYPOTHETICAL 45.8 KDA PROTEIN SLL0103.	*S.sp.*	21.0	241 – 407	17 – 188	3	4.00E-39
◇	▷	AAB84957.1	magnesium chelatase subunit ChlI	*M.thermoautotrophicum*	25.0	151 – 370	324 – 535	3	5.00E-22
◇	▷	CAA30479.1	Not given	*M.musculus*	20.0	247 – 416	132 – 309	4	2.00E-44
◇	▷	AAA59544.1	Not given	*H.sapiens*	16.0	223 – 412	107 – 305	4	3.00E-42
◇	▷	AAB24821.1	leukocyte integrin alpha chain	*H.sapiens*	16.0	223 – 412	107 – 305	4	3.00E-42
◇	▷	1JLM	I-DOMAIN FROM INTEGRIN CR3, MN2+ BOUND	*H.sapiens*	16.0	263 – 412	1 – 158	4	2.00E-41
◇	▷	1IDO	I-DOMAIN FROM INTEGRIN CR3, MG2+ BOUND	*H.sapiens*	16.0	263 – 412	1 – 158	4	2.00E-41
◇	▷	1BHO:1	MAC-1 I DOMAIN MAGNESIUM COMPLEX	*H.sapiens*	16.0	264 – 412	2 – 158	4	3.00E-41
◇	▷	1IDN:2	MAC-1 I DOMAIN METAL FREE	*H.sapiens*	16.0	264 – 412	2 – 158	4	3.00E-41
◇	▷	1IDN:1	MAC-1 I DOMAIN METAL FREE	*H.sapiens*	16.0	264 – 412	2 – 158	4	3.00E-41
◇	▷	1BHO:2	MAC-1 I DOMAIN MAGNESIUM COMPLEX	*H.sapiens*	16.0	264 – 412	2 – 158	4	3.00E-41
◇	▷	1BHQ:2	MAC-1 I DOMAIN CADMIUM COMPLEX	*H.sapiens*	16.0	264 – 412	2 – 158	4	3.00E-41
◇	▷	1BHQ:1	MAC-1 I DOMAIN CADMIUM COMPLEX	*H.sapiens*	16.0	264 – 412	2 – 158	4	3.00E-41
◇	▷	O22437	MAGNESIUM-CHELATASE SUBUNIT CHLD PRECURSO	*P.sativum*	19.0	162 – 417	468 – 701	4	7.00E-41
◇	▷	CAB69268.1	Not given	*unidentified*	18.0	162 – 417	473 – 705	4	2.00E-40
◇	▷	AAB16869.1	CD11b	*S.scrofa*	17.0	254 – 417	6 – 177	4	8.00E-40
◇	▷	CAB58179.1	Mg-protoporphyrin IX	*H.vulgare*	17.0	162 – 410	90 – 311	4	2.00E-38
◇	▷	AAD52031.1	magnesium chelatase subunit CHLD precursor	*A.thaliana*	17.0	162 – 423	441 – 695	4	5.00E-38
◇	▷	BAA16787.1	Mg chelatase subunit	*S.sp.*	18.0	199 – 367	415 – 588	4	6.00E-38
◇	▷	AAB86281.1	conserved protein	*M.thermoautotrophicum*	13.0	154 – 393	60 – 303	4	7.00E-37
◇	▷	CAA77537.1	561 aa (60 kD) Mg chelatase subunit	*R.capsulatus*	21.0	168 – 367	277 – 484	4	1.00E-36
◇	▷	AAF22895.1	T27G7.20	*A.thaliana*	17.0	162 – 423	491 – 747	4	2.00E-36
◇	▷	AAF21241.1	alpha D integrin	*R.norvegicus*	14.0	239 – 412	126 – 307	4	2.00E-35
◇	▷	AAF23492.1	leukocyte adhesion glycoprotein p150,95 alpha integrin su	*M.musculus*	13.0	231 – 412	118 – 307	4	2.00E-35
◇	▷	AAB38547.1	leukointegrin alpha d chain	*H.sapiens*	13.0	239 – 410	124 – 303	4	3.00E-35

▷ GenBank ▷ SWISS-PROT ▷ PDB

AlEye output (December 13, 2000 3:07 PM)

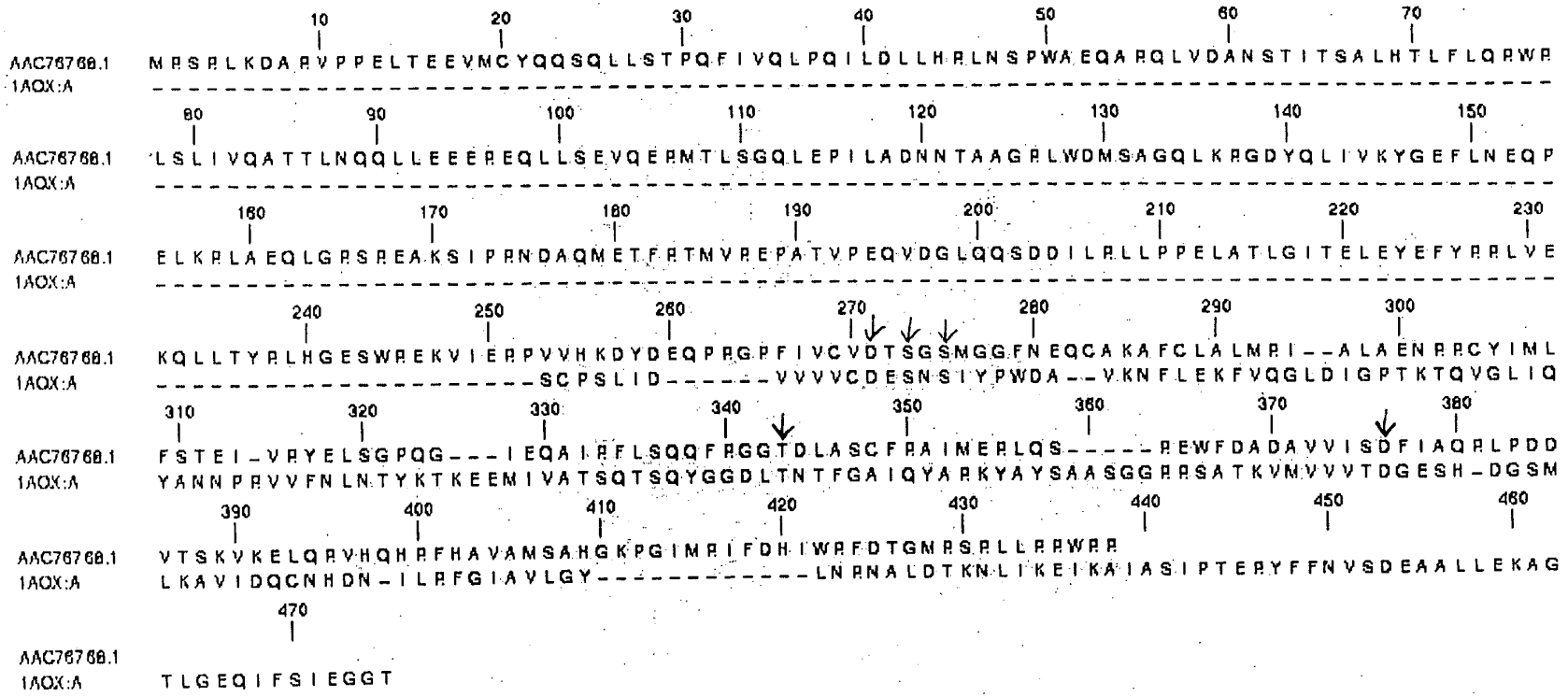


Figure 16A

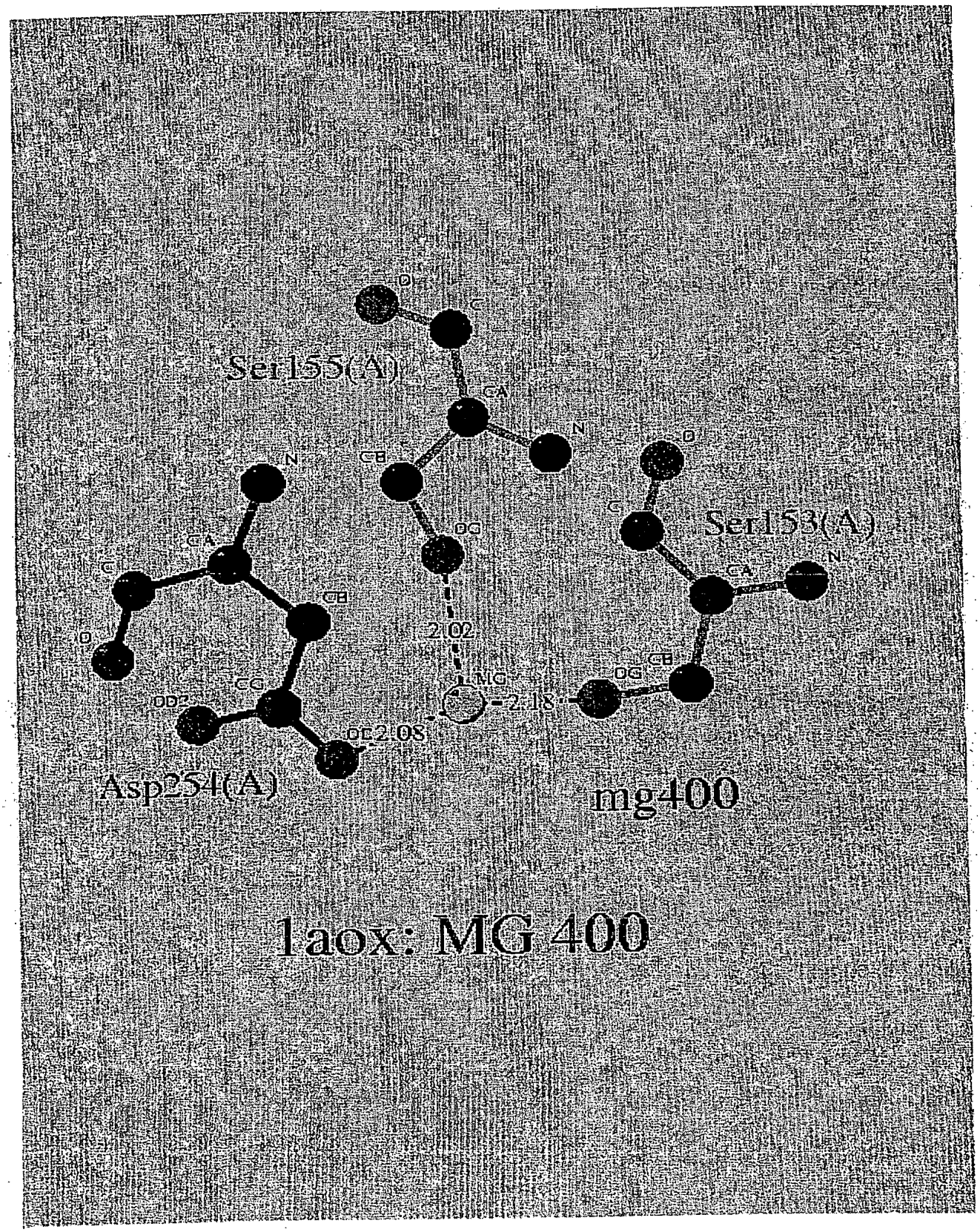


Figure 16B

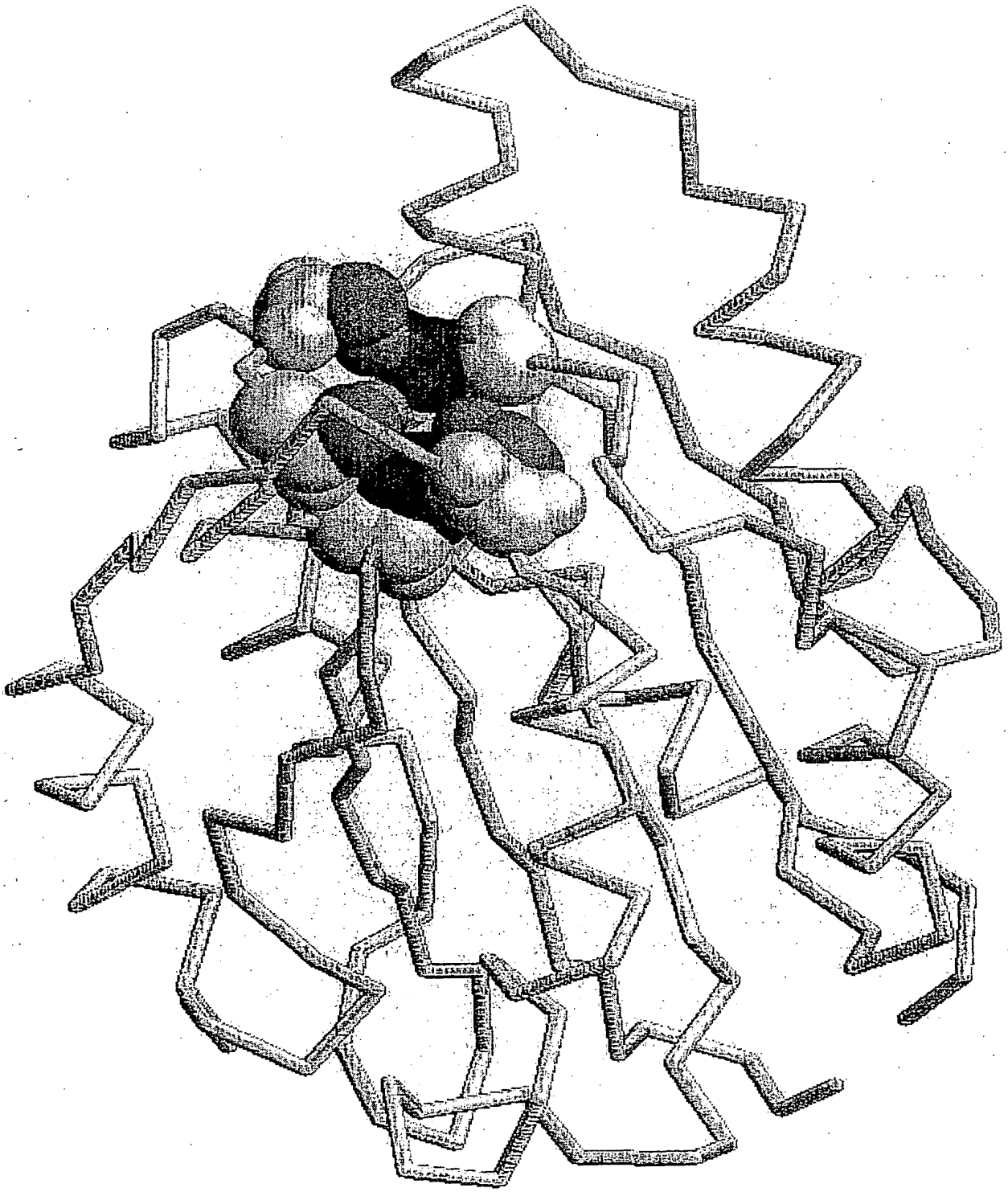
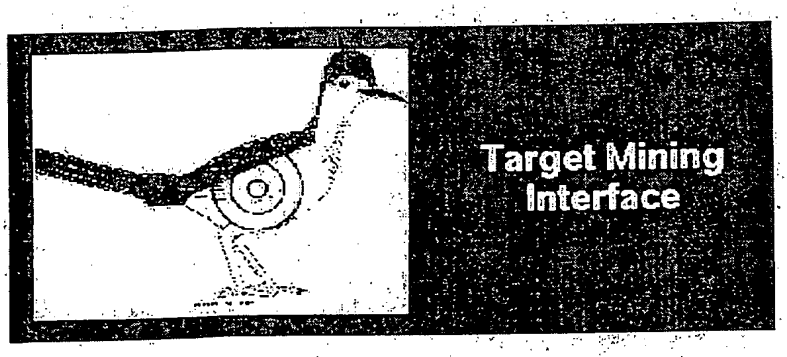


Figure 17

File Edit View Go Communicator Help

Bookmarks Location http://victoria.inpharmatica.co.uk/~volker/BPD3target.html



inpharmatica

Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): 1JLM and chain (e.g. B):

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release: DEvF9 - BPD3

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

None

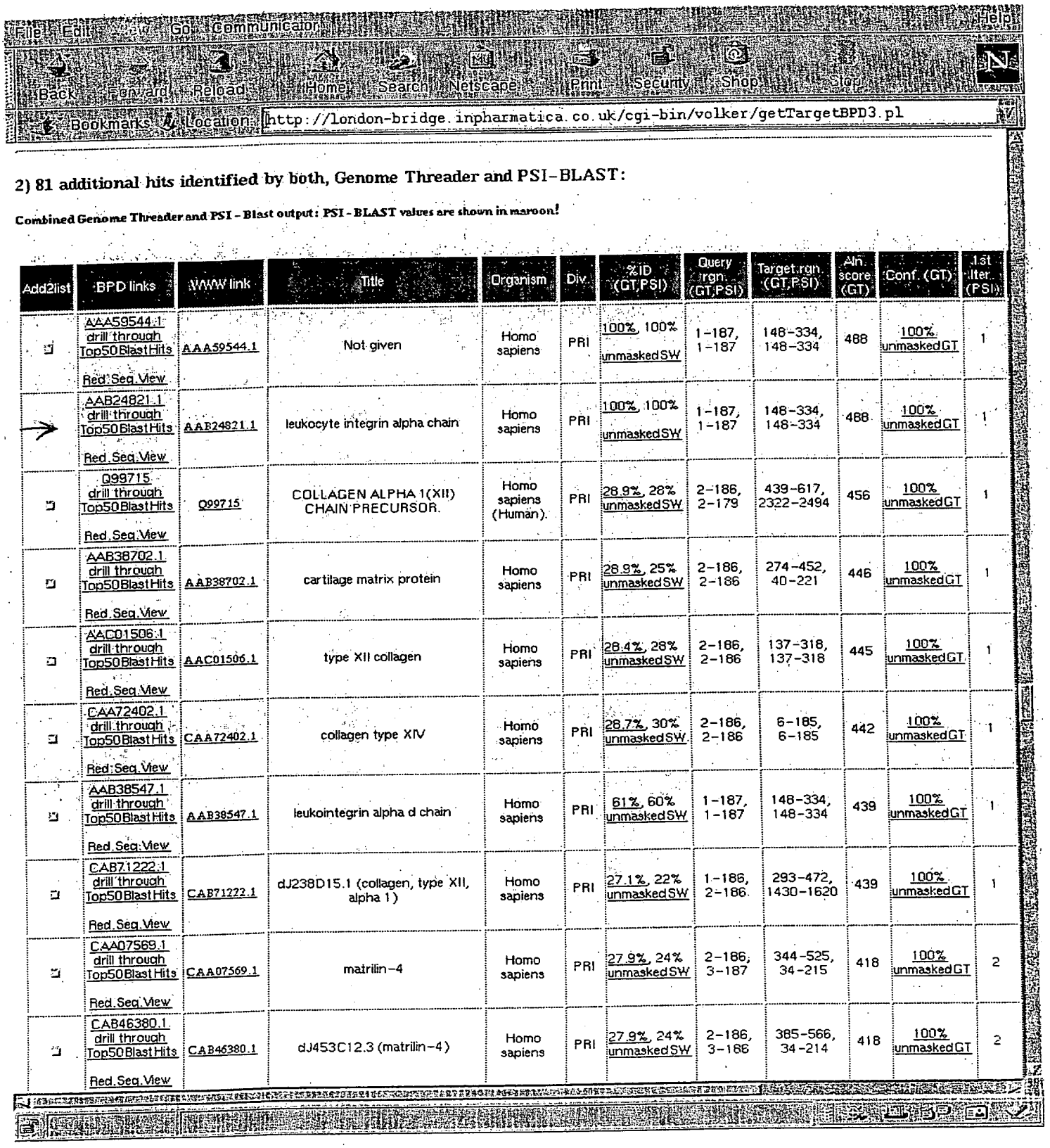
If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

- Filter for the following SPECIES:

Homo sapiens | Rattus norvegicus (Rat) | Mus musculus (Mouse) | Danio rerio (zebrafish)

100%

Figure 18A



2) 81 additional hits identified by both, Genome Threader and PSI-BLAST:

Combined Genome Threader and PSI - Blast output: PSI - BLAST values are shown in maroon!

Add2list	BPD links	WWW link	Title	Organism	Div.	%ID (GT,PSI)	Query rgn. (GT,PSI)	Target rgn. (GT,PSI)	Aln. score (GT)	Conf. (GT)	1st iter. (PSI)
	AAA59544.1 drill through Top50BlastHits Red.Seq.View	AAA59544.1	Not given	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	488	100% unmaskedGT	1
→	AAB24821.1 drill through Top50BlastHits Red.Seq.View	AAB24821.1	leukocyte integrin alpha chain	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	488	100% unmaskedGT	1
	Q99715 drill through Top50BlastHits Red.Seq.View	Q99715	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR.	Homo sapiens (Human).	PRI	28.9%, 28% unmaskedSW	2-186, 2-179	439-617, 2322-2494	456	100% unmaskedGT	1
	AAB38702.1 drill through Top50BlastHits Red.Seq.View	AAB38702.1	cartilage matrix protein	Homo sapiens	PRI	28.9%, 25% unmaskedSW	2-186, 2-186	274-452, 40-221	446	100% unmaskedGT	1
	AAC01506.1 drill through Top50BlastHits Red.Seq.View	AAC01506.1	type XII collagen	Homo sapiens	PRI	28.4%, 28% unmaskedSW	2-186, 2-186	137-318, 137-318	445	100% unmaskedGT	1
	CAA72402.1 drill through Top50BlastHits Red.Seq.View	CAA72402.1	collagen type XIV	Homo sapiens	PRI	28.7%, 30% unmaskedSW	2-186, 2-186	6-185, 6-185	442	100% unmaskedGT	1
	AAB38547.1 drill through Top50BlastHits Red.Seq.View	AAB38547.1	leukointegrin alpha d chain	Homo sapiens	PRI	61%, 60% unmaskedSW	1-187, 1-187	148-334, 148-334	439	100% unmaskedGT	1
	CAB71222.1 drill through Top50BlastHits Red.Seq.View	CAB71222.1	dJ238D15.1 (collagen, type XII, alpha 1)	Homo sapiens	PRI	27.1%, 22% unmaskedSW	1-186, 2-186	293-472, 1430-1620	439	100% unmaskedGT	1
	CAA07569.1 drill through Top50BlastHits Red.Seq.View	CAA07569.1	matrilin-4	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-187	344-525, 34-215	418	100% unmaskedGT	2
	CAB46380.1 drill through Top50BlastHits Red.Seq.View	CAB46380.1	dJ453C12.3 (matrilin-4)	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-186	385-566, 34-214	418	100% unmaskedGT	2

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location: http://london-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl

	Red.Seq.View	drill through Top50BlastHits	BAA31611.1	KIAA0636 protein	Homo sapiens	PRI	20.2% unmaskedSW	73-185	390-503	66	unmaskedGT reverse Hit
	Red.Seq.View	CAB52192.1 drill through Top50BlastHits	CAB52192.1	G7c protein	Homo sapiens	PRI	14.1% unmaskedSW	6-88	20-105	76	89.66% unmaskedGT
	Red.Seq.View	BAA75899.1 drill through Top50BlastHits	BAA75899.1	N-copine	Homo sapiens	PRI	26.8% unmaskedSW	73-143	405-474	73	89.66% unmaskedGT
	Red.Seq.View	P26012 drill through Top50BlastHits	P26012	INTEGRIN BETA-8 PRECURSOR.	Homo sapiens (Human).	PRI	13.1% unmaskedSW	2-115	147-289	78	87.73% unmaskedGT
	Red.Seq.View	AAD21820.1 drill through Top50BlastHits	AAD21820.1	NG37	Homo sapiens	PRI	14.1% unmaskedSW	6-88	318-403	76	85.59% unmaskedGT
	Red.Seq.View	CAA65775.1 drill through Top50BlastHits	CAA65775.1	Sec23 protein	Homo sapiens	PRI	19.7% unmaskedSW	7-141	134-275	74	84.45% unmaskedGT
	Red.Seq.View	AAC15920.1 drill through Top50BlastHits	AAC15920.1	copine I	Homo sapiens	PRI	20% unmaskedSW	73-185	384-497	67	80.79% unmaskedGT reverse Hit
	Red.Seq.View	AAC63290.1 drill through Top50BlastHits	AAC63290.1	polymerase	Homo sapiens	PRI	18.8% unmaskedSW	84-185	442-552	55	74.2% unmaskedGT
	Red.Seq.View	CAB87610.1 drill through Top50BlastHits	CAB87610.1	dJ309K20.3 (Copine I (similar to KIAA0636))	Homo sapiens	PRI	25% unmaskedSW	76-143	363-429	67	72.84% unmaskedGT
	Red.Seq.View	AAA35532.1 drill through Top50BlastHits	AAA35532.1	Not given	Homo sapiens	PRI	11.1% unmaskedSW	6-145	373-504	56	71.48% unmaskedGT
→	Red.Seq.View	P10155 drill through Top50BlastHits	P10155	60 KDA RO PROTEIN (60 KDA RIBONUCLEOPROTEIN RO)(RORNP)(SJOGREN SYNDROME TYPE A ANTIGEN (SS-A)).	Homo sapiens (Human).	PRI	11.1% unmaskedSW	6-145	373-504	56	70.11% unmaskedGT
	Red.Seq.View	AAC15863.1 drill through Top50BlastHits	AAC15863.1	hexokinase I	Homo sapiens	PRI	18.5% unmaskedSW	22-122	749-851	75	68.76% unmaskedGT reverse Hit
	Red.Seq.View	BAA92672.1 drill through Top50BlastHits	BAA92672.1	KIAA1434 protein	Homo sapiens	PRI	20.8% unmaskedSW	124-186	576-647	67	67.41% unmaskedGT
	Red.Seq.View	CAA10335.1 drill through Top50BlastHits	CAA10335.1	Sec24B protein	Homo sapiens	PRI	9% unmaskedSW	6-91	682-801	61	67.41% unmaskedGT
	Red.Seq.View	AAD43766.1 drill through Top50BlastHits	AAD43766.1	protocadherin gamma A11 short form protein	Homo sapiens	PRI	15.4% unmaskedSW	30-185	475-643	66	66.06% unmaskedGT
	Red.Seq.View	AAD43714.1 drill through Top50BlastHits	AAD43714.1	protocadherin gamma A11	Homo sapiens	PRI	15.4% unmaskedSW	30-185	475-643	66	63.4% unmaskedGT
	Red.Seq.View	CAA08933.1 drill through Top50BlastHits	CAA08933.1	annexin 31 (annexin XXXI)	Homo sapiens	PRI	10.2% unmaskedSW	20-184	123-308	58	63.4% unmaskedGT
	Red.Seq.View	AAA52646.1 drill through Top50BlastHits	AAA52646.1	hexokinase I	Homo sapiens	PRI	18.5% unmaskedSW	22-122	750-852	73	62.08% unmaskedGT reverse Hit
	Red.Seq.View	AAA59185.1 drill through Top50BlastHits	AAA59185.1	integrin beta-7 subunit	Homo sapiens	PRI	16.4% unmaskedSW	2-55	152-204	67	59.48% unmaskedGT
	Red.Seq.View	AAB65421.1 drill through Top50BlastHits	AAB65421.1	integrin variant beta4E	Homo sapiens	PRI	24.3% unmaskedSW	2-36	130-166	73	56.92% unmaskedGT

Figure 18B

Figure 18C

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location: http://london-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl

649 PSI-BLAST matches were identified:

[illegible]

626 out of these 649 PSI-BLAST matches were identified using 'positive iterations':

[illegible]

23 out of these 649 PSI-BLAST matches were identified using 'negative iterations':

[illegible]

A2) Genome Threader Matches:

2925 matches found by Genome Threading:

[illegible]

Figure 19

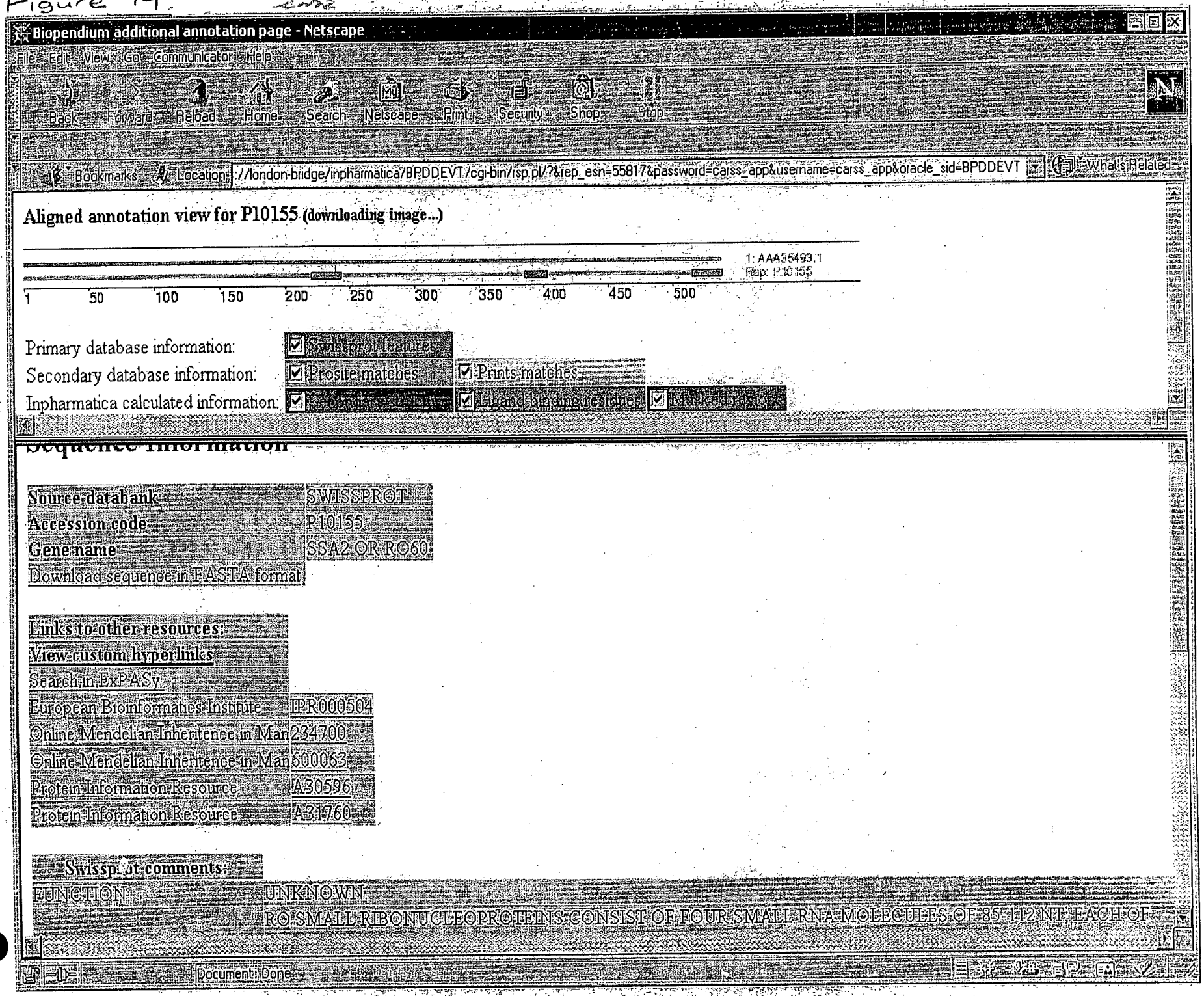
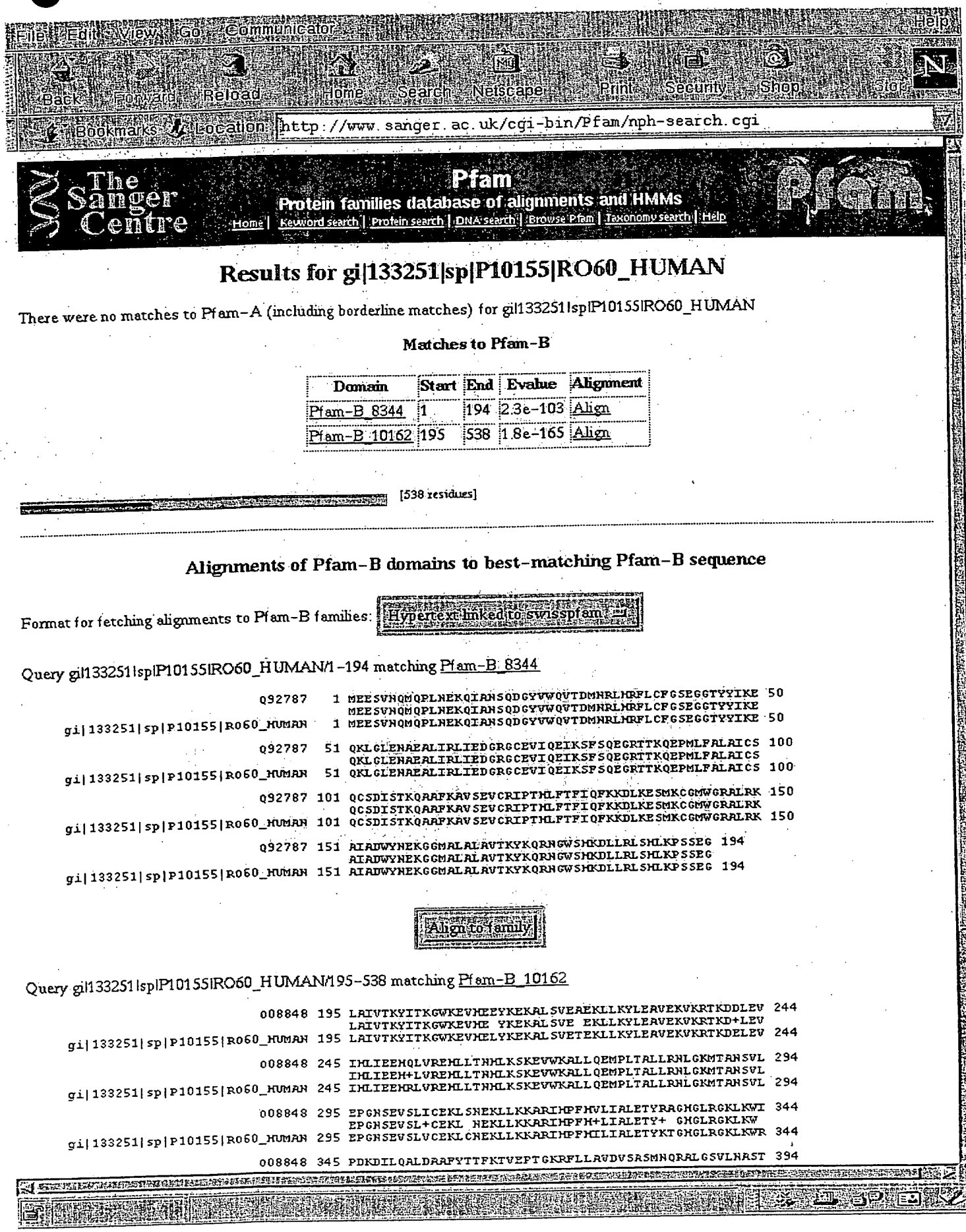

Biopendium additional annotation page - Netscape
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: //london-bridge/inpharmatica/BPDDEVT/cgi-bin/isp.pl/?&rep_esn=55817&password=carss_app&username=carss_app&oracle_sid=BPDDEVT What's Related
Aligned annotation view for P10155 (downloading image...)
1: AAA35499.1
Rep: P10155
1 50 100 150 200 250 300 350 400 450 500
Primary database information:
Swissprot features
Secondary database information:
Prosite matches
Prints matches
Inpharmatica calculated information:
Ligand binding residues
Sequence information
Source databank SWISSPROT
Accession code P10155
Gene name SSA2 OR RO60
Download sequence in FASTA format
Links to other resources:
View custom hyperlinks
Search in ExPASy
European Bioinformatics Institute IPR000504
Online Mendelian Inheritence in Man 234700
Online Mendelian Inheritence in Man 600063
Protein Information Resource A30596
Protein Information Resource A31760
Swissprot comments:
FUNCTION UNKNOWN
RO SMALL RIBONUCLEOPROTEINS CONSIST OF FOUR SMALL RNA MOLECULES OF 85-112 NT, EACH OF
Document: Done

Figure 20



Results for gi|133251|sp|P10155|RO60_HUMAN

There were no matches to Pfam-A (including borderline matches) for gi|133251|sp|P10155|RO60_HUMAN

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_8344	1	194	2.3e-103	Align
Pfam-B_10162	195	538	1.8e-165	Align

[538 residues]

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families: Hypertext linked to swisspfam

Query gi|133251|sp|P10155|RO60_HUMAN/1-194 matching Pfam-B_8344

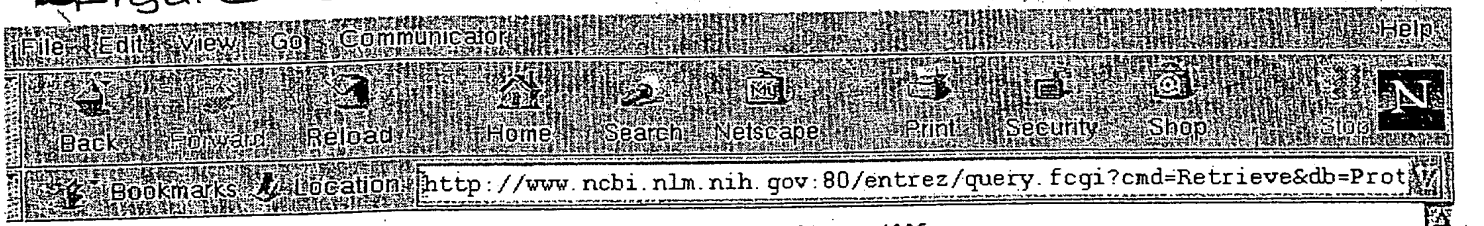
```
                    Q92787    1 MEESVNQMQPLNEKQIANSQDGYVWQVTDMNRLMRFLCFGSEGGTYYIKE  50
                              MEESVNQMQPLNEKQIANSQDGYVWQVTDMNRLMRFLCFGSEGGTYYIKE
gi|133251|sp|P10155|RO60_HUMAN    1 MEESVNQMQPLNEKQIANSQDGYVWQVTDMNRLMRFLCFGSEGGTYYIKE  50

                    Q92787   51 QKLGLENAEALIRLIEDGRGCEVIQEIKSFSQEGRTTKQEPMLFALAICS 100
                              QKLGLENAEALIRLIEDGRGCEVIQEIKSFSQEGRTTKQEPMLFALAICS
gi|133251|sp|P10155|RO60_HUMAN   51 QKLGLENAEALIRLIEDGRGCEVIQEIKSFSQEGRTTKQEPMLFALAICS 100

                    Q92787  101 QCSDISTKQAAFKAVSEVCRIPTHLFTFIQFKKDLKESMKCGMWGRALRK 150
                              QCSDISTKQAAFKAVSEVCRIPTHLFTFIQFKKDLKESMKCGMWGRALRK
gi|133251|sp|P10155|RO60_HUMAN  101 QCSDISTKQAAFKAVSEVCRIPTHLFTFIQFKKDLKESMKCGMWGRALRK 150

                    Q92787  151 AIADWYNEKGGMALALAVTKYKQRNGWSMKDLLRLSMLKPSSEG 194
                              AIADWYNEKGGMALALAVTKYKQRNGWSMKDLLRLSMLKPSSEG
gi|133251|sp|P10155|RO60_HUMAN  151 AIADWYNEKGGMALALAVTKYKQRNGWSMKDLLRLSMLKPSSEG 194
```

Align to family

Query gi|133251|sp|P10155|RO60_HUMAN/195-538 matching Pfam-B_10162

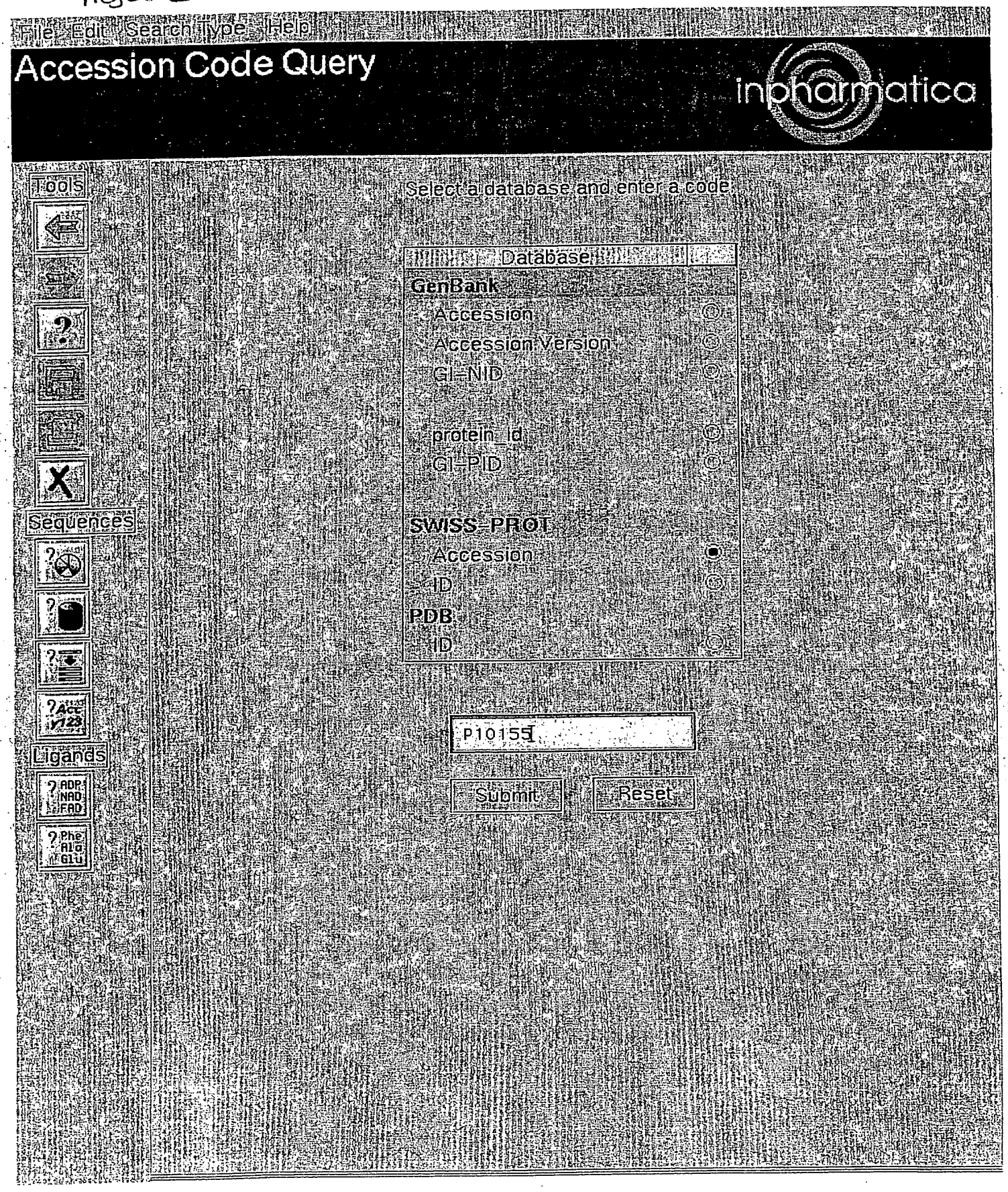
```
                    O08848  195 LAIVTKYITKGWKEVHEEYKEKALSVEAEKLLKYLEAVEKVKRTKDDLEV 244
                              LAIVTKYITKGWKEVHE YKEKALSVE EKLLKYLEAVEKVKRTKD+LEV
gi|133251|sp|P10155|RO60_HUMAN  195 LAIVTKYITKGWKEVHELYKEKALSVETEKLLKYLEAVEKVKRTKDELEV 244

                    O08848  245 IHLIEEHQLVREHLLTNHLKSKEVWKALLQEMPLTALLRNLGKMTANSVL 294
                              IHLIEEH+LVREHLLTNHLKSKEVWKALLQEMPLTALLRNLGKMTANSVL
gi|133251|sp|P10155|RO60_HUMAN  245 IHLIEEHRLVREHLLTNHLKSKEVWKALLQEMPLTALLRNLGKMTANSVL 294

                    O08848  295 EPGNSEVSLICEKLSNEKLLKKARIHPFHVLIALETYRAGHGLRGKLKWI 344
                              EPGNSEVSL+CEKL NEKLLKKARIHPFH+LIALETY+ GHGLRGKLKW
gi|133251|sp|P10155|RO60_HUMAN  295 EPGNSEVSLVCEKLCNEKLLKKARIHPFHILIALETYKTGHGLRGKLKWR 344

                    O08848  345 PDKDILQALDRAFYTTFKTVEPTGKRFLLRVDVSRSMNQRALGSVLNAST 394
```

Figure 21




```
LOCUS       RO60_HUMAN     538 aa                      PRI       01-FEB-1996
DEFINITION  60 KD RO PROTEIN (60 KD RIBONUCLEOPROTEIN RO) (RORNP) (SJOGREN
            SYNDROME TYPE A ANTIGEN (SS-A)).
ACCESSION   P10155
PID         g133251
VERSION     P10155  GI:133251
DBSOURCE    swissprot: locus RO60_HUMAN, accession P10155;
            class: standard.
            created: Mar 1, 1989.
            sequence updated: Mar 1, 1989.
            annotation updated: Feb 1, 1996.
            xrefs: gi: gi: 177782, gi: gi: 177783, gi: gi: 387656, gi: gi:
            387657, gi: gi: 86722, gi: gi: 107626
            xrefs (non-sequence databases): MIM 600063, MIM 234700, PROSITE
            PS00030
KEYWORDS    Ribonucleoprotein; RNA-binding; Systemic lupus erythematosus;
            Antigen.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 538)
  AUTHORS   Deutscher,S.L., Harley,J.B. and Keene,J.D.
  TITLE     Molecular analysis of the 60-kDa human Ro ribonucleoprotein
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85 (24), 9479-9483 (1988)
  MEDLINE   89071722
  REMARK    SEQUENCE FROM N.A.
REFERENCE   2  (residues 1 to 538)
  AUTHORS   Ben-Chetrit,E., Gandy,B.J., Tan,E.M. and Sullivan,K.F.
  TITLE     Isolation and characterization of a cDNA clone encoding the 60-kD
            component of the human SS-A/Ro ribonucleoprotein autoantigen
  JOURNAL   J. Clin. Invest. 83 (4), 1284-1292 (1989)
  MEDLINE   89198084
  REMARK    SEQUENCE FROM N.A.
COMMENT     ---------------------------------------------------------------
            This SWISS-PROT entry is copyright. It is produced through a
            collaboration between the Swiss Institute of Bioinformatics and
            the EMBL outstation - the European Bioinformatics Institute.
            The original entry is available from http://www.expasy.ch/sprot
            and http://www.ebi.ac.uk/sprot
            ---------------------------------------------------------------.
            [FUNCTION] UNKNOWN.
            [SUBUNIT] RO SMALL RIBONUCLEOPROTEINS CONSIST OF FOUR SMALL RNA
            MOLECULES OF 85-112 NT, EACH OF WHICH IS COMPLEXED WITH A 60 KD
            PROTEIN. RO RNPS MAY ALSO CONTAIN AN ADDITIONAL 52 KD PROTEIN.
            [SUBCELLULAR LOCATION] CYTOPLASMIC.
            [DISEASE] SERA FROM PATIENTS WITH SYSTEMIC LUPUS ERYTHEMATOSUS
            OFTEN CONTAIN ANTIBODIES THAT REACT WITH THE NORMAL CELLULAR RO
            PROTEIN AS IF THESE ANTIGEN WAS FOREIGN.
            [SIMILARITY] CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
            [SIMILARITY] STRONG, TO XENOPUS 60 KD RO PROTEIN.
FEATURES                Location/Qualifiers
     source             1..538
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        1..538
     Protein            1..538
                        /product="60 KD RO PROTEIN"
     Region             93..98
                        /region_name="Domain"
                        /note="RNA-BINDING (RNP2) (BY SIMILARITY)."
     Region             124..131
                        /region_name="Domain"
                        /note="RNA-BINDING (RNP1) (BY SIMILARITY)."
     Region             239
                        /region_name="Conflict"
                        /note="K -> R (IN REF. 2)."
     Region             515..538
                        /region_name="Conflict"
                        /note="GMLDMCGFDTGALDVIRNFTLDMI -> ALQNTLLNKSF (IN REF.
                        2)."
ORIGIN
        1 meesvnqmqp lnekqiansq dgyvwqvtdm nrllhrflcfg seggtpyike qklglenaea
       61 lirliedgrg ceviqeiksf sqegrttkqe pmlfalaics qcsdistkqa afkavsevcr
      121 ipthlftfiq fkkdlkesmk cgmwgralrk aiadwynekg gmalalavtk ykqrngwshk
      181 dllrlshlkp sseglaivtk yitkgwkevh elykekalsv eteklikyle avekvkrtkd
      241 elevihliee hrlvrehllt nhllkskevwk allqemplta llrnlgkmta nsvlepgnse
      301 vslvceklcn eklikkarih pfhililalet yktghglrgk lkwrpdeeil kaldaafykt
      361 fktveptgkr fllavdvsas mnqrvlgsil nastvaaamc mvvtrtekds yvvafsdemv
      421 pcpvttdmtl qqvlmamsqi paggtdcslp miwaqktntp advfivftdn etfaggvhpa
      481 ialreyrkkm dipaklivcg mtsngftiad pddrgmldmc gfdtgaldvi rnftldmi
//
```

Figure 22A


File Edit Search Type Help
Accession Code Query
inpharmatica
Tools
Sequences
Ligands
Select a database and enter a code:
Database
GenBank
Accession
Accession.Version
GI-NID
protein_id
GI-PID
SWISS-PROT
Accession
ID
PDB
ID
P10155
Submit
Reset

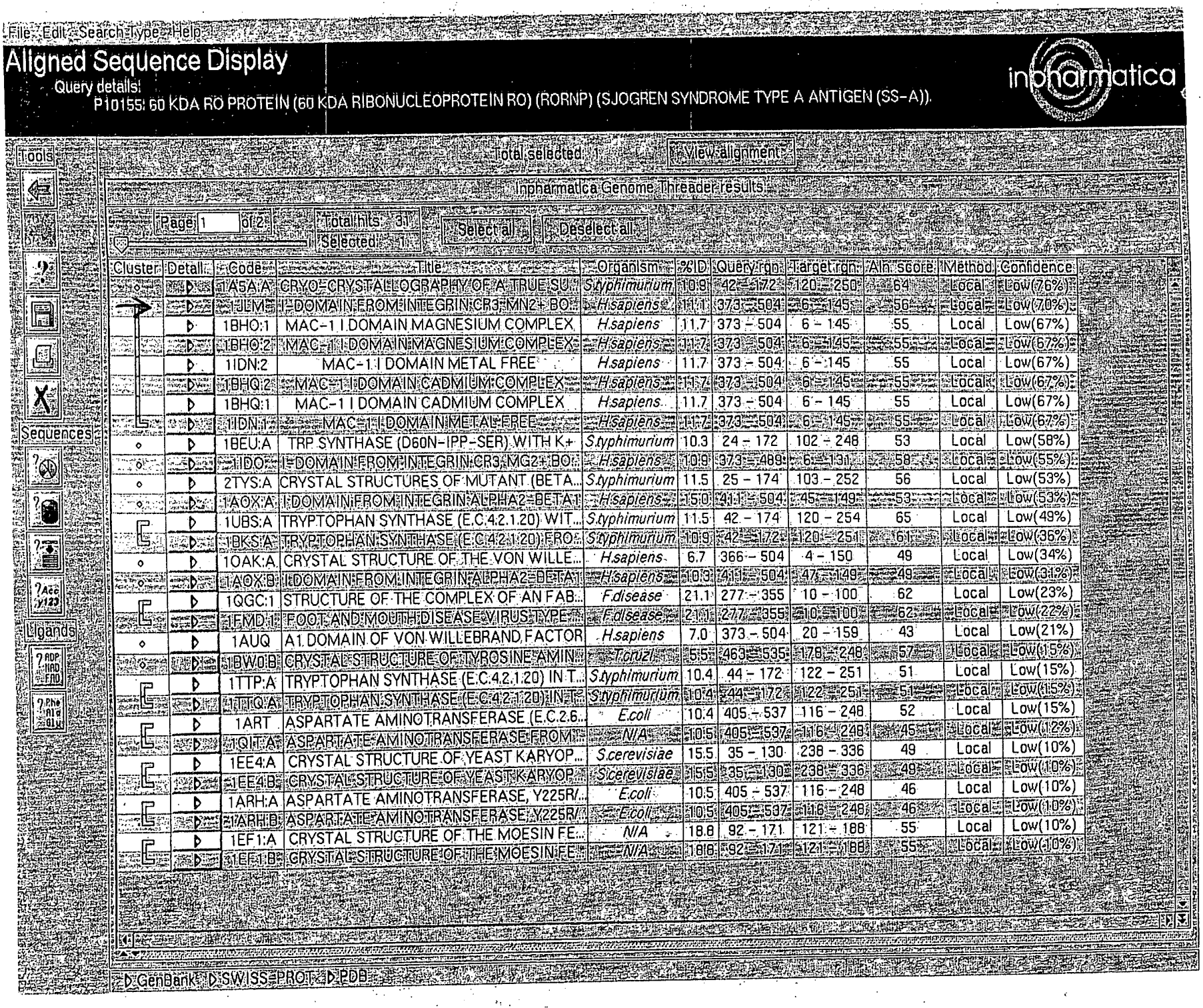

File Edit Search Type Help
Aligned Sequence Display
Query details:
P10155: 60 KDA RO PROTEIN (60 KDA RIBONUCLEOPROTEIN RO) (RORNP) (SJOGREN SYNDROME TYPE A ANTIGEN (SS-A)).
inpharmatica
Tools
Sequences
Ligands
Total selected: 1
View alignment
Inpharmatica Genome Threader results
Page 1 of 2
Total hits: 31
Selected: 1
Select all
Deselect all
Cluster | Detail | Code | Title | Organism | %ID | Query rgn | Target rgn | Aln score | Method | Confidence
1A5A:A | CRYO-CRYSTALLOGRAPHY OF A TRUE SU... | S.typhimurium | 10.9 | 42 - 172 | 120 - 250 | 64 | Local | Low(76%)
1JLM:1 | I-DOMAIN FROM INTEGRIN CR3 MN2+ BO... | H.sapiens | 11.1 | 373 - 504 | 6 - 145 | 56 | Local | Low(70%)
1BHO:1 | MAC-1 I DOMAIN MAGNESIUM COMPLEX | H.sapiens | 11.7 | 373 - 504 | 6 - 145 | 55 | Local | Low(67%)
1BHO:2 | MAC-1 I DOMAIN MAGNESIUM COMPLEX | H.sapiens | 11.7 | 373 - 504 | 6 - 145 | 55 | Local | Low(67%)
1IDN:2 | MAC-1 I DOMAIN METAL FREE | H.sapiens | 11.7 | 373 - 504 | 6 - 145 | 55 | Local | Low(67%)
1BHQ:2 | MAC-1 I DOMAIN CADMIUM COMPLEX | H.sapiens | 11.7 | 373 - 504 | 6 - 145 | 55 | Local | Low(67%)
1BHQ:1 | MAC-1 I DOMAIN CADMIUM COMPLEX | H.sapiens | 11.7 | 373 - 504 | 6 - 145 | 55 | Local | Low(67%)
1IDN:1 | MAC-1 I DOMAIN METAL-FREE | H.sapiens | 11.7 | 373 - 504 | 6 - 145 | 55 | Local | Low(67%)
1BEU:A | TRP SYNTHASE (D60N-IPP-SER) WITH K+ | S.typhimurium | 10.3 | 24 - 172 | 102 - 248 | 53 | Local | Low(58%)
1IDO:1 | I-DOMAIN FROM INTEGRIN CR3 MG2+ BO... | H.sapiens | 10.9 | 373 - 489 | 6 - 131 | 58 | Local | Low(55%)
2TYS:A | CRYSTAL STRUCTURES OF MUTANT (BETA... | S.typhimurium | 11.5 | 25 - 174 | 103 - 252 | 56 | Local | Low(53%)
1AOX:A | I-DOMAIN FROM INTEGRIN ALPHA2-BETA1 | H.sapiens | 5.0 | 411 - 504 | 45 - 149 | 53 | Local | Low(53%)
1UBS:A | TRYPTOPHAN SYNTHASE (E.C.4.2.1.20) WIT... | S.typhimurium | 11.5 | 42 - 174 | 120 - 254 | 65 | Local | Low(49%)
1BKS:A | TRYPTOPHAN SYNTHASE (E.C.4.2.1.20) FRO... | S.typhimurium | 10.9 | 42 - 172 | 120 - 251 | 61 | Local | Low(36%)
1OAK:A | CRYSTAL STRUCTURE OF THE VON WILLE... | H.sapiens | 6.7 | 366 - 504 | 4 - 150 | 49 | Local | Low(34%)
1AOX:B | I-DOMAIN FROM INTEGRIN ALPHA2-BETA1 | H.sapiens | 10.3 | 411 - 504 | 47 - 149 | 49 | Local | Low(31%)
1QGC:1 | STRUCTURE OF THE COMPLEX OF AN FAB... | F.disease | 21.1 | 277 - 355 | 10 - 100 | 62 | Local | Low(23%)
1FMD:1 | FOOT AND MOUTH DISEASE VIRUS TYPE ... | F.disease | 21.1 | 277 - 355 | 10 - 100 | 62 | Local | Low(22%)
1AUQ | A1 DOMAIN OF VON WILLEBRAND FACTOR | H.sapiens | 7.0 | 373 - 504 | 20 - 159 | 43 | Local | Low(21%)
1BW0:B | CRYSTAL STRUCTURE OF TYROSINE AMIN... | T.cruzi | 5.5 | 463 - 535 | 178 - 248 | 57 | Local | Low(15%)
1TTP:A | TRYPTOPHAN SYNTHASE (E.C.4.2.1.20) IN T... | S.typhimurium | 10.4 | 44 - 172 | 122 - 251 | 51 | Local | Low(15%)
1TTQ:A | TRYPTOPHAN SYNTHASE (E.C.4.2.1.20) IN T... | S.typhimurium | 10.4 | 44 - 172 | 122 - 251 | 51 | Local | Low(15%)
1ART | ASPARTATE AMINOTRANSFERASE (E.C.2.6... | E.coli | 10.4 | 405 - 537 | 116 - 248 | 52 | Local | Low(15%)
1QIT:A | ASPARTATE AMINOTRANSFERASE FROM ... | N/A | 10.5 | 405 - 537 | 116 - 248 | 45 | Local | Low(12%)
1EE4:A | CRYSTAL STRUCTURE OF YEAST KARYOP... | S.cerevisiae | 15.5 | 35 - 130 | 238 - 336 | 49 | Local | Low(10%)
1EE4:B | CRYSTAL STRUCTURE OF YEAST KARYOP... | S.cerevisiae | 15.5 | 35 - 130 | 238 - 336 | 49 | Local | Low(10%)
1ARH:A | ASPARTATE AMINOTRANSFERASE, Y225R/... | E.coli | 10.5 | 405 - 537 | 116 - 248 | 46 | Local | Low(10%)
1ARH:B | ASPARTATE AMINOTRANSFERASE, Y225R/... | E.coli | 10.5 | 405 - 537 | 116 - 248 | 46 | Local | Low(10%)
1EF1:A | CRYSTAL STRUCTURE OF THE MOESIN FE... | N/A | 18.8 | 92 - 171 | 121 - 188 | 55 | Local | Low(10%)
1EF1:B | CRYSTAL STRUCTURE OF THE MOESIN FE... | N/A | 18.8 | 92 - 171 | 121 - 188 | 55 | Local | Low(10%)
GenBank SWISS-PROT PDB

Figure 22B

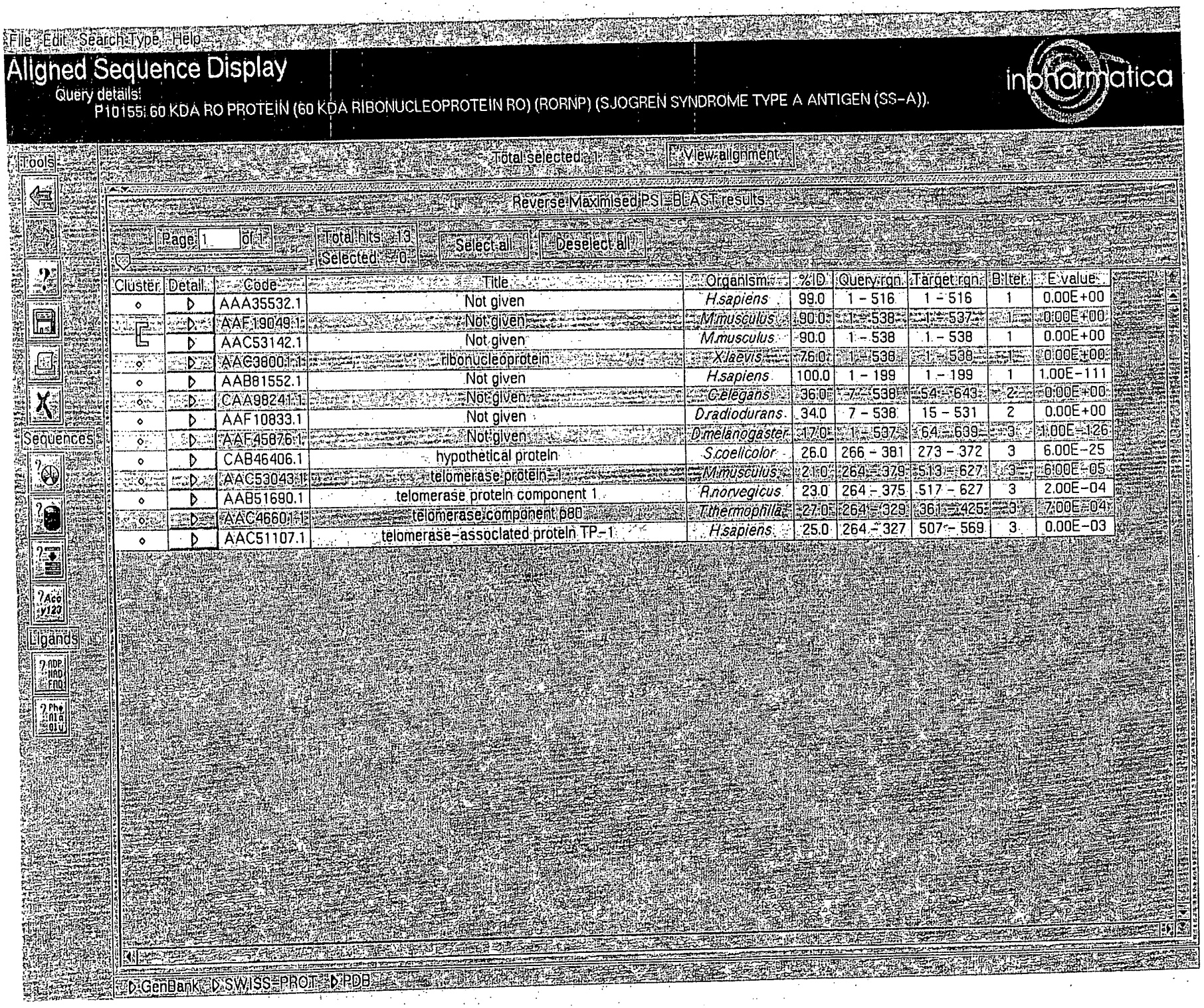

File Edit Search Type Help
Aligned Sequence Display
Query details:
P10155 60 KDA RO PROTEIN (60 KDA RIBONUCLEOPROTEIN RO) (RORNP) (SJOGREN SYNDROME TYPE A ANTIGEN (SS-A)).
inpharmatica
Tools
Sequences
Ligands
Total selected: 1
View alignment
Reverse Maximised PSI-BLAST results
Page 1 of 1
Total hits: 13
Selected: 0
Select all
Deselect all
Cluster Detail Code Title Organism %ID Query rgn. Target rgn. Iter. E value
AAA35532.1 Not given H.sapiens 99.0 1 - 516 1 - 516 1 0.00E+00
AAF19049.1 Not given M.musculus 90.0 1 - 538 1 - 537 1 0.00E+00
AAC53142.1 Not given M.musculus 90.0 1 - 538 1 - 538 1 0.00E+00
AAG38001.1 ribonucleoprotein X.laevis 76.0 1 - 536 1 - 538 1 0.00E+00
AAB81552.1 Not given H.sapiens 100.0 1 - 199 1 - 199 1 1.00E-111
CAA98241.1 Not given C.elegans 36.0 7 - 538 54 - 643 2 0.00E+00
AAF10833.1 Not given D.radiodurans 34.0 7 - 538 15 - 531 2 0.00E+00
AAF45876.1 Not given D.melanogaster 17.0 1 - 537 64 - 639 3 1.00E-126
CAB46406.1 hypothetical protein S.coelicolor 26.0 266 - 381 273 - 372 3 6.00E-25
AAC53043.1 telomerase protein 1 M.musculus 21.0 264 - 379 513 - 627 3 6.00E-05
AAB51690.1 telomerase protein component 1 R.norvegicus 23.0 264 - 375 517 - 627 3 2.00E-04
AAC46601.1 telomerase component p80 T.thermophila 27.0 264 - 329 361 - 425 3 7.00E-04
AAC51107.1 telomerase-associated protein TP-1 H.sapiens 25.0 264 - 327 507 - 569 3 0.00E-03
GenBank SWISS-PROT PDB

Figure 22C

Figure 23

AlEye output (January 2, 2001 1:29 PM)

```
        10        20        30        40        50        60        70        80
P10155  MEESVNQMQPLNEKQIANSQDGYVWQVTDMNRLHRFLCFGSEGGTYYIKEQKLGLENAEALIRLIEDGRGCEVIQEIKSF
1JJM    --------------------------------------------------------------------------------

        90        100       110       120       130       140       150       160
P10155  SQEGRTTKQEPMLFALAICSQCSDISTKQAAFKAVSEVCRIPTHLFTFIQFKKDLKESMKCGMWGRALRKAIADWYNEKG
1JJM    --------------------------------------------------------------------------------

        170       180       190       200       210       220       230       240
P10155  GMALALAVTKYKQRNGWSHKDLLRLSHLKPSSEGLAIVTKYITKGWKEVHELYKEKALSVETEKLLKYLEAVEKVKRTKD
1JJM    --------------------------------------------------------------------------------

        250       260       270       280       290       300       310       320
P10155  ELEVIHLIEEHRLVREHLLTNHLKSKEVWKALLQEMPLTALLRNLGKMTANSVLEPGNSEVSLVCEKLCNEKLLKKARIH
1JJM    --------------------------------------------------------------------------------

        330       340       350       360       370       380       390       400
P10155  PFHILIALETYKTGHGLRGKLKWRPDEEILKALDAAFYKTFKTVEPTGKRFL--L--AVDVSASMNQRVLGSILNASTVA
1JJM    ---------------------------------------------------DSDIAFLLDGSGSIIPHDFRRMKEFVSTV

        410       420       430       440       450       460       470       480
P10155  A-----AMCMVVTRTEKDSYVVAFSDEMVPCPVTTDMTLQQVLMAMSQI-PAGGTDCSLPMIWAQKTNT---PADVFIVF
1JJM    MEQLKKSKTLFSLMQYSEEFRIHFTFKEFQNNPNPRSLVKPITQLLGRTHTATGIRKVVRELFNITNGARKNAFKILVVI

        490       500       510       520       530       540       550       560
P10155  TDNETFAGGVHPAIALREYRKKMDIPAKLIVCGMTSNGF-------TIADPDDRG-MLDMCGFDTGALDVIRNFTLDMI
1JJM    TDGEKFGDPLGYEDVIPE-ADREGVIRYVIGVGDAFRSEKSRQELNTIASKPPRDHVFQVNNFE--ALKTIQNQLREKIF

P10155
1JJM    A
```

Figure 24A

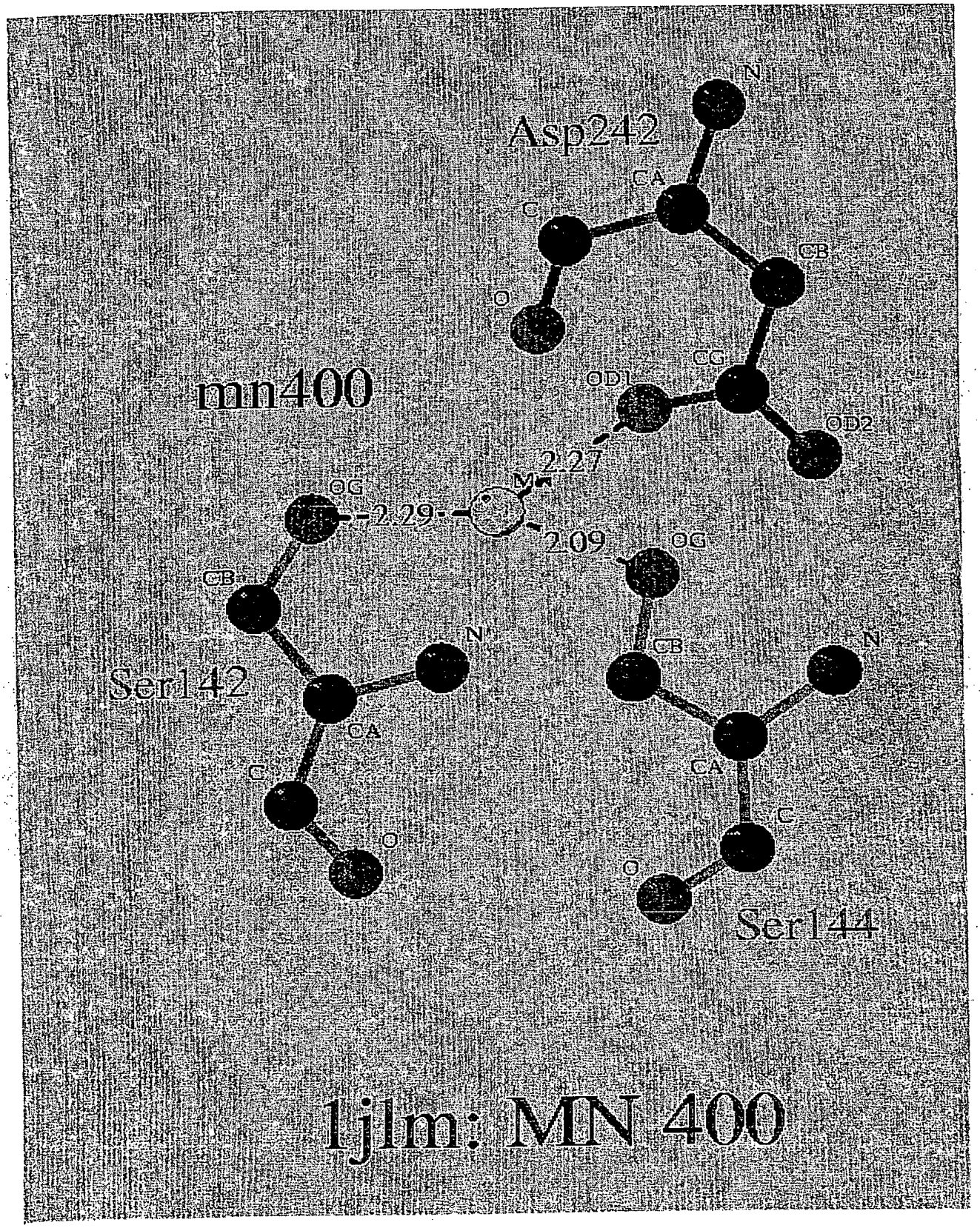
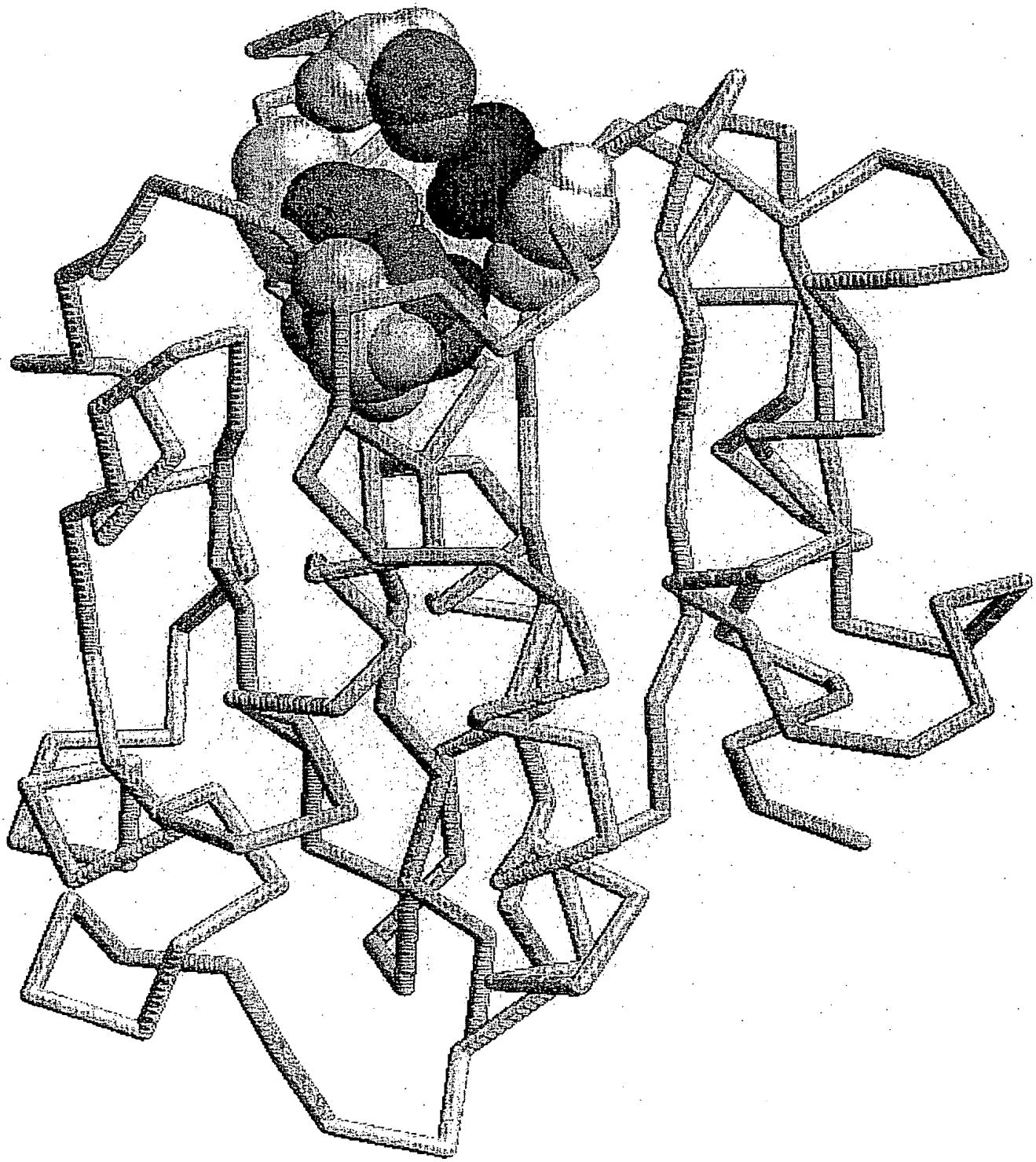

Asp242
N
CA
C
CB
O
CG
OD1
OD2
mn400
2.27
Mn
2.29
2.09
OG
CB
N
Ser142
CA
C
O
OG
CB
N
CA
C
O
Ser144
1jlm: MN 400

Figure 24B



AlEye output (January 4, 2001 3:18 PM)

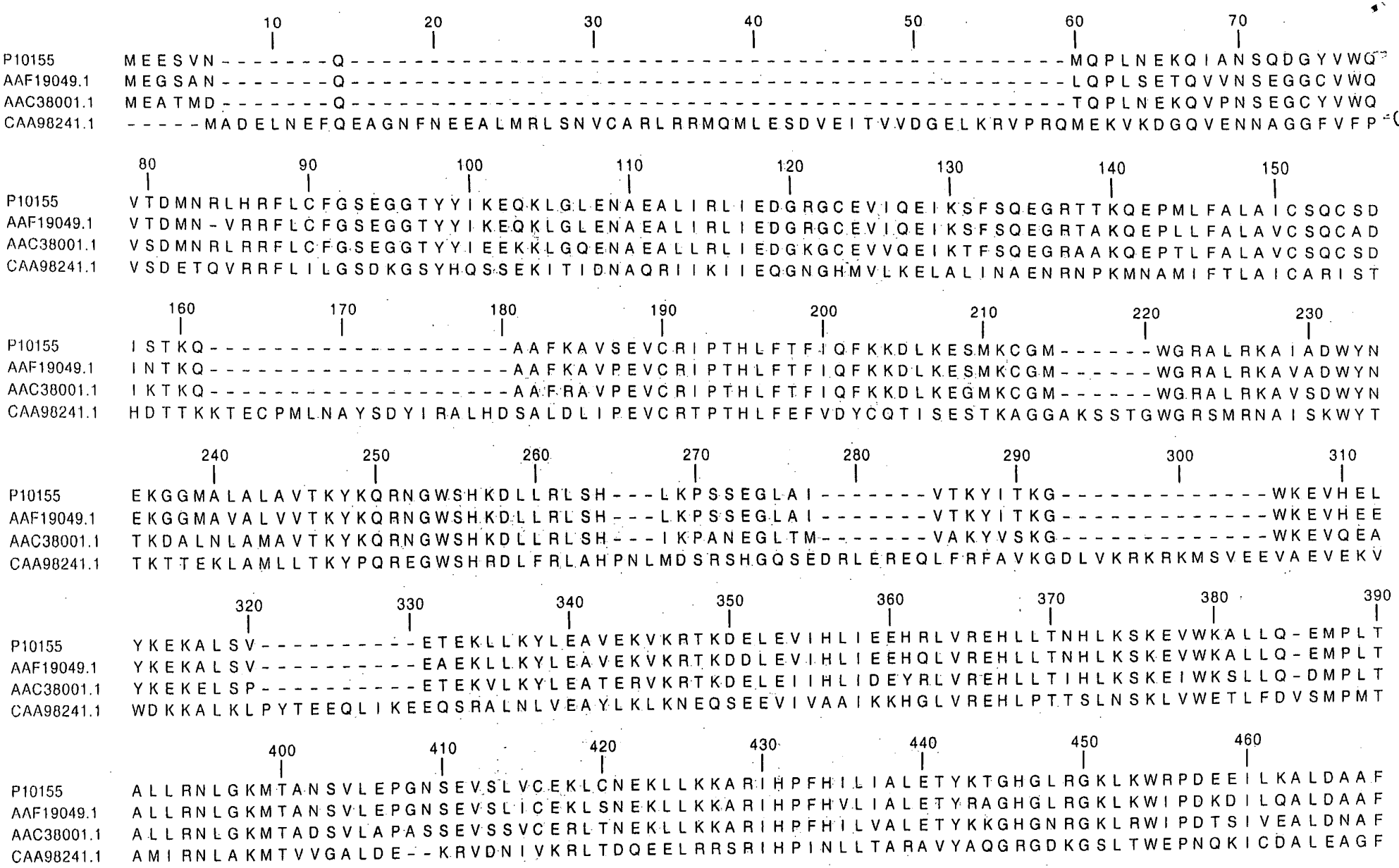


Figure 2B

AlEye output (January 4, 2001 3:18 PM)

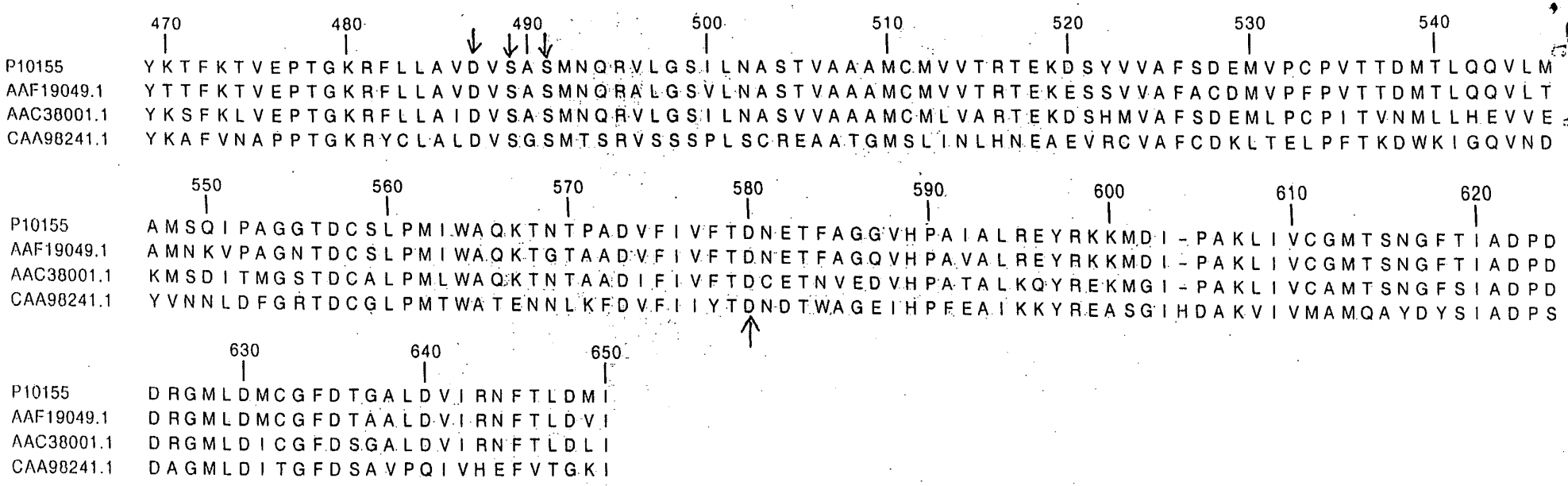


Figure 2 cont'd.